



Europäisches Patentamt  
European Patent Office  
Office européen des brevets



(11) **EP 0 786 519 A2**

(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:  
30.07.1997 Bulletin 1997/31

(51) Int Cl<sup>6</sup> **C12N 15/00**

(21) Application number **97100117.7**

(22) Date of filing **07.01.1997**

(84) Designated Contracting States:  
**AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC  
NL PT SE**  
Designated Extension States:  
**AL LT LV RO SI**

(30) Priority: **05.01.1996 US 9861**

(71) Applicant: **HUMAN GENOME SCIENCES, INC.**  
**Rockville, MD 20850-3338 (US)**

(72) Inventors:  
• **Kunsch, Charles A.**  
**Gaithersburg, Maryland 20882 (US)**

- **Choi, Gil H.**  
**Rockville, Maryland 20850 (US)**
- **Barash, Steven C.**  
**Rockville, Maryland 20850 (US)**
- **Dillon, Patrick J.**  
**Gaithersburg, Maryland 20879 (US)**
- **Fannon, Michael R.**  
**Silver Spring, Maryland 20906 (US)**
- **Rosen, Craig A.**  
**Laytonsville, Maryland 20882 (US)**

(74) Representative **VOSSIUS & PARTNER**  
**Postfach 86 07 67**  
**81634 München (DE)**

(54) **Staphylococcus aureus polynucleotides and sequences**

(57) The present invention provides polynucleotide sequences of the genome of *Staphylococcus aureus*, polypeptide sequences encoded by the polynucleotide sequences, corresponding polynucleotides and polypeptides, vectors and hosts comprising the polynu-

cleotides, and assays and other uses thereof. The present invention further provides polynucleotide and polypeptide sequence information stored on computer readable media, and computer-based systems and methods which facilitate its use.

**EP 0 786 519 A2**

## Description

The present invention relates to the field of molecular biology. In particular, it relates to, among other things, nucleotide sequences of *Staphylococcus aureus*, contigs, ORFs, fragments, probes, primers and related polynucleotides thereof, peptides and polypeptides encoded by the sequences, and uses of the polynucleotides and sequences thereof, such as in fermentation, polypeptide production, assays and pharmaceutical development, among others.

The genus *Staphylococcus* includes at least 20 distinct species. (For a review see Novick, R. P., *The Staphylococcus as a Molecular Genetic System*, Chapter 1, pgs. 1-37 in *MOLECULAR BIOLOGY OF THE STAPHYLOCOCCI*, R. Novick, Ed., VCH Publishers, New York (1990)). Species differ from one another by 80% or more, by hybridization kinetics, whereas strains within a species are at least 90% identical by the same measure.

The species *Staphylococcus aureus*, a gram-positive, facultatively aerobic, clump-forming cocci, is among the most important etiological agents of bacterial infection in humans, as discussed briefly below.

## Human Health and *S. Aureus*

*Staphylococcus aureus* is a ubiquitous pathogen. (See, for instance, Mims *et al.*, *MEDICAL MICROBIOLOGY*, Mosby-Year Book Europe Limited, London, UK (1993)). It is an etiological agent of a variety of conditions, ranging in severity from mild to fatal. A few of the more common conditions caused by *S. aureus* infection are burns, cellulitis, eyelid infections, food poisoning, joint infections, neonatal conjunctivitis, osteomyelitis, skin infections, surgical wound infection, scalded skin syndrome and toxic shock syndrome, some of which are described further below.

### Burns

Burn wounds generally are sterile initially. However, they generally compromise physical and immune barriers to infection, cause loss of fluid and electrolytes and result in local or general physiological dysfunction. After cooling, contact with viable bacteria results in mixed colonization at the injury site. Infection may be restricted to the non-viable debris on the burn surface ("eschar"), it may progress into full skin infection and invade viable tissue below the eschar and it may reach below the skin, enter the lymphatic and blood circulation and develop into septicaemia. *S. aureus* is among the most important pathogens typically found in burn wound infections. It can destroy granulation tissue and produce severe septicaemia.

### Cellulitis

Cellulitis, an acute infection of the skin that expands from a typically superficial origin to spread below the cutaneous layer, most commonly is caused by *S. aureus* in conjunction with *S. pyogenes*. Cellulitis can lead to systemic infection. In fact, cellulitis can be one aspect of synergistic bacterial gangrene. This condition typically is caused by a mixture of *S. aureus* and microaerophilic streptococci. It causes necrosis and treatment is limited to excision of the necrotic tissue. The condition often is fatal.

### Eyelid infections

*S. aureus* is the cause of styes and of sticky eye\* in neonates, among other eye infections. Typically such infections are limited to the surface of the eye, and may occasionally penetrate the surface with more severe consequences.

### Food poisoning

Some strains of *S. aureus* produce one or more of five serologically distinct, heat and acid stable enterotoxins that are not destroyed by digestive process of the stomach and small intestine (enterotoxins A-E). Ingestion of the toxin, in sufficient quantities, typically results in severe vomiting, but not diarrhoea. The effect does not require viable bacteria. Although the toxins are known, their mechanism of action is not understood.

### Joint infections

*S. aureus* infects bone joints causing diseases such as osteomyelitis.

### Osteomyelitis

*S. aureus* is the most common causative agent of haematogenous osteomyelitis. The disease tends to occur in



children and adolescents more than adults and it is associated with non-penetrating injuries to bones. Infection typically occurs in the long end of growing bone, hence its occurrence in physically immature populations. Most often, infection is localized in the vicinity of sprouting capillary loops adjacent to epiphyseal growth plates in the end of long, growing bones

#### *Skin infections*

*S. aureus* is the most common pathogen of such minor skin infections as abscesses and boils. Such infections often are resolved by normal host response mechanisms, but they also can develop into severe internal infections. Recurrent infections of the nasal passages plague nasal carriers of *S. aureus*

#### *Surgical Wound Infections*

Surgical wounds often penetrate far into the body. Infection of such wound thus poses a grave risk to the patient. *S. aureus* is the most important causative agent of infections in surgical wounds. *S. aureus* is unusually adept at invading surgical wounds: sutured wounds can be infected by far fewer *S. aureus* cells than are necessary to cause infection in normal skin. Invasion of surgical wound can lead to severe *S. aureus* septicaemia. Invasion of the blood stream by *S. aureus* can lead to seeding and infection of internal organs, particularly heart valves and bone, causing systemic diseases, such as endocarditis and osteomyelitis.

#### *Scalded Skin Syndrome*

*S. aureus* is responsible for "scalded skin syndrome" (also called toxic epidermal necrosis, Ritter's disease and Lyell's disease). This disease occurs in older children, typically in outbreaks caused by flowering of *S. aureus* strains produce exfoliation (also called scalded skin syndrome toxin). Although the bacteria initially may infect only a minor lesion, the toxin destroys intercellular connections, spreads epidermal layers and allows the infection to penetrate the outer layer of the skin, producing the desquamation that typifies the disease. Shedding of the outer layer of skin generally reveals normal skin below, but fluid lost in the process can produce severe injury in young children if it is not treated properly.

#### *Toxic Shock Syndrome*

Toxic shock syndrome is caused by strains of *S. aureus* that produce the so-called toxic shock syndrome toxin. The disease can be caused by *S. aureus* infection at any site, but it is too often erroneously viewed exclusively as a disease solely of women who use tampons. The disease involves toxemia and septicaemia, and can be fatal.

#### *Nocosomal Infections*

In the 1984 National Nosocomial Infection Surveillance Study ("NNIS") *S. aureus* was the most prevalent agent of surgical wound infections in many hospital services, including medicine, surgery, obstetrics, pediatrics and newborns.

#### *Resistance to drugs of S. aureus strains*

Prior to the introduction of penicillin the prognosis for patients seriously infected with *S. aureus* was unfavorable. Following the introduction of penicillin in the early 1940s even the worst *S. aureus* infections generally could be treated successfully. The emergence of penicillin-resistant strains of *S. aureus* did not take long, however. Most strains of *S. aureus* encountered in hospital infections today do not respond to penicillin, although, fortunately, this is not the case for *S. aureus* encountered in community infections.

It is well known now that penicillin-resistant strains of *S. aureus* produce a lactamase which converts penicillin to penicilloic acid, and thereby destroys antibiotic activity. Furthermore, the lactamase gene often is propagated episomally, typically on a plasmid, and often is only one of several genes on an episomal element that, together, confer multidrug resistance.

Methicillins, introduced in the 1960s, largely overcame the problem of penicillin resistance in *S. aureus*. These compounds conserve the portions of penicillin responsible for antibiotic activity and modify or alter other portions that make penicillin a good substrate for inactivating lactamases. However, methicillin resistance has emerged in *S. aureus*, along with resistance to many other antibiotics effective against this organism, including aminoglycosides, tetracycline, chloramphenicol, macrolides and lincosamides. In fact, methicillin-resistant strains of *S. aureus* generally are multiply drug resistant.

The molecular genetics of most types of drug resistance in *S. aureus* has been elucidated (See Lyon *et al.*, *Microbiology Reviews* 51: 88-134 (1987)). Generally, resistance is mediated by plasmids, as noted above regarding penicillin resistance, however, several stable forms of drug resistance have been observed that apparently involve integration of a resistance element into the *S. aureus* genome itself.

Thus far each new antibiotic gives rise to resistance strains, strains emerge that are resistance to multiple drugs and increasingly persistent forms of resistance begin to emerge. Drug resistance of *S. aureus* infections already poses significant treatment difficulties, which are likely to get much worse unless new therapeutic agents are developed.

### Molecular Genetics of *Staphylococcus Aureus*

Despite its importance in, among other things, human disease, relatively little is known about the genome of this organism.

Most genetic studies of *S. aureus* have been carried out using the strain NCTC8325, which contains prophages psi11, psi12 and psi13, and the UV-cured derivative of this strain, 8325-4 (also referred to as RN450), which is free of the prophages.

These studies revealed that the *S. aureus* genome, like that of other *staphylococci*, consists of one circular, covalently closed, double-stranded DNA and a collection of so-called variable accessory genetic elements, such as prophages, plasmids, transposons and the like.

Physical characterization of the genome has not been carried out in any detail. Pattee *et al.* published a low resolution and incomplete genetic and physical map of the chromosome of *S. aureus* strain NCTC 8325. (Pattee *et al.* Genetic and Physical Mapping of Chromosome of *Staphylococcus aureus* NCTC 8325, Chapter 11, pgs. 163-169 in: MOLECULAR BIOLOGY OF THE STAPHYLOCOCCI, R.P. Novick, Ed., VCH Publishers, New York, (1990). The genetic map largely was produced by mapping insertions of Tn551 and Tn4001, which, respectively, confer erythromycin and gentamicin resistance, and by analysis of SmaI-digested DNA by Pulsed Field Gel Electrophoresis ("PFGE").

The map was of low resolution; even estimating the physical size of the genome was difficult, according to the investigators. The size of the largest SmaI chromosome fragment, for instance, was too large for accurate sizing by PFGE. To estimate its size, additional restriction sites had to be introduced into the chromosome using a transposon containing a SmaI recognition sequence.

In sum, most physical characteristics and almost all of the genes of *Staphylococcus aureus* are unknown. Among the few genes that have been identified, most have not been physically mapped or characterized in detail. Only a very few genes of this organism have been sequenced. (See, for instance Thornsberry, J., *Antimicrobial Chemotherapy* 21 Suppl. C: 9-16 (1988), current versions of GENBANK and other nucleic acid databases, and references that relate to the genome of *S. aureus* such as those set out elsewhere herein.)

It is clear that the etiology of diseases mediated or exacerbated by *S. aureus* infection involves the programmed expression of *S. aureus* genes, and that characterizing the genes and their patterns of expression would add dramatically to our understanding of the organism and its host interactions. Knowledge of *S. aureus* genes and genomic organization would dramatically improve understanding of disease etiology and lead to improved and new ways of preventing, ameliorating, arresting and reversing diseases. Moreover, characterized genes and genomic fragments of *S. aureus* would provide reagents for, among other things, detecting, characterizing and controlling *S. aureus* infections. There is a need therefore to characterize the genome of *S. aureus* and for polynucleotides and sequences of this organism.

The present invention is based on the sequencing of fragments of the *Staphylococcus aureus* genome. The primary nucleotide sequences which were generated are provided in SEQ ID NOS: 1-5,191.

The present invention provides the nucleotide sequence of several thousand contigs of the *Staphylococcus aureus* genome, which are listed in tables below and set out in the Sequence Listing submitted herewith, and representative fragments thereof, in a form which can be readily used, analyzed, and interpreted by a skilled artisan. In one embodiment, the present invention is provided as contiguous strings of primary sequence information corresponding to the nucleotide sequences depicted in SEQ ID NOS 1-5,191.

The present invention further provides nucleotide sequences which are at least 95%, preferably 99% and most preferably 99.9%, identical to the nucleotide sequences of SEQ ID NOS: 1-5,191.

The nucleotide sequence of SEQ ID NOS 1-5,191, a representative fragment thereof, or a nucleotide sequence which is at least 95%, preferably 99% and most preferably 99.9%, identical to the nucleotide sequence of SEQ ID NOS 1-5,191 may be provided in a variety of mediums to facilitate its use. In one application of this embodiment, the sequences of the present invention are recorded on computer readable media. Such media includes, but is not limited to magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape, optical storage media such as CD-ROM, electrical storage media such as RAM and ROM, and hybrids of these categories such as magnetic/optical storage media.

The present invention further provides systems, particularly computer-based systems which contain the sequence

information herein described stored in a data storage means. Such systems are designed to identify commercially important fragments of the *Staphylococcus aureus* genome.

Another embodiment of the present invention is directed to fragments, preferably isolated fragments, of the *Staphylococcus aureus* genome having particular structural or functional attributes. Such fragments of the *Staphylococcus aureus* genome of the present invention include, but are not limited to, fragments which encode peptides, hereinafter referred to as open reading frames or ORFs,\* fragments which modulate the expression of an operably linked ORF, hereinafter referred to as expression modulating fragments or EMFs,\* and fragments which can be used to diagnose the presence of *Staphylococcus aureus* in a sample, hereinafter referred to as diagnostic fragments or "DFs".

Each of the ORFs in fragments of the *Staphylococcus aureus* genome disclosed in Tables 1-3, and the EMFs found 5' to the ORFs, can be used in numerous ways as polynucleotide reagents. For instance, the sequences can be used as diagnostic probes or amplification primers for detecting or determining the presence of a specific microbe in a sample, to selectively control gene expression in a host and in the production of polypeptides, such as polypeptides encoded by ORFs of the present invention, particular those polypeptides that have a pharmacological activity.

The present invention further includes recombinant constructs comprising one or more fragments of the *Staphylococcus aureus* genome of the present invention. The recombinant constructs of the present invention comprise vectors, such as a plasmid or viral vector, into which a fragment of the *Staphylococcus aureus* has been inserted.

The present invention further provides host cells containing any of the isolated fragments of the *Staphylococcus aureus* genome of the present invention. The host cells can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic cell, such as a yeast cell, or a procaryotic cell such as a bacterial cell.

The present invention is further directed to polypeptides and proteins, preferably isolated polypeptides and proteins, encoded by ORFs of the present invention. A variety of methods, well known to those of skill in the art, routinely may be utilized to obtain any of the polypeptides and proteins of the present invention. For instance, polypeptides and proteins of the present invention having relatively short, simple amino acid sequences readily can be synthesized using commercially available automated peptide synthesizers. Polypeptides and proteins of the present invention also may be purified from bacterial cells which naturally produce the protein. Yet another alternative is to purify polypeptide and proteins of the present invention from cells which have been altered to express them.

The invention further provides polypeptides, preferably isolated polypeptides, comprising *Staphylococcus aureus* epitopes and vaccine compositions comprising such polypeptides. Also provided are methods for vaccinating an individual against *Staphylococcus aureus* infection.

The invention further provides methods of obtaining homologs of the fragments of the *Staphylococcus aureus* genome of the present invention and homologs of the proteins encoded by the ORFs of the present invention. Specifically, by using the nucleotide and amino acid sequences disclosed herein as a probe or as primers, and techniques such as PCR cloning and colony/plaque hybridization, one skilled in the art can obtain homologs.

The invention further provides antibodies which selectively bind polypeptides and proteins of the present invention. Such antibodies include both monoclonal and polyclonal antibodies.

The invention further provides hybridomas which produce the above-described antibodies. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

The present invention further provides methods of identifying test samples derived from cells which express one of the ORFs of the present invention, or a homolog thereof. Such methods comprise incubating a test sample with one or more of the antibodies of the present invention, or one or more of the DFs or antigens of the present invention, under conditions which allow a skilled artisan to determine if the sample contains the ORF or product produced therefrom.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the above-described assays.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the antibodies, antigens, or one of the DFs of the present invention, and (b) one or more other containers comprising one or more of the following wash reagents, reagents capable of detecting presence of bound antibodies, antigens or hybridized DFs.

Using the isolated proteins of the present invention, the present invention further provides methods of obtaining and identifying agents capable of binding to a polypeptide or protein encoded by one of the ORFs of the present invention. Specifically, such agents include, as further described below, antibodies, peptides, carbohydrates, pharmaceutical agents and the like. Such methods comprise steps of: (a) contacting an agent with an isolated protein encoded by one of the ORFs of the present invention; and (b) determining whether the agent binds to said protein.

The present genomic sequences of *Staphylococcus aureus* will be of great value to all laboratories working with this organism and for a variety of commercial purposes. Many fragments of the *Staphylococcus aureus* genome will be immediately identified by similarity searches against GenBank or protein databases and will be of immediate value to *Staphylococcus aureus* researchers and for immediate commercial value for the production of proteins or to control gene expression.

The methodology and technology for elucidating extensive genomic sequences of bacterial and other genomes

has and will greatly enhance the ability to analyze and understand chromosomal organization. In particular, sequenced contigs and genomes will provide the models for developing tools for the analysis of chromosome structure and function, including the ability to identify genes within large segments of genomic DNA, the structure, position, and spacing of regulatory elements, the identification of genes with potential industrial applications, and the ability to do comparative genomic and molecular phylogeny.

**FIGURE 1** is a block diagram of a computer system (102) that can be used to implement computer-based systems of present invention.

**FIGURE 2** is a schematic diagram depicting the data flow and computer programs used to collect, assemble, edit and annotate the contigs of the *Staphylococcus aureus* genome of the present invention. Both Macintosh and Unix platforms are used to handle the AB 373 and 377 sequence data files, largely as described in Kerlavage *et al.*, *Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System Sciences*, 585, IEEE Computer Society Press, Washington D.C. (1993). Factura (AB) is a Macintosh program designed for automatic vector sequence removal and end-trimming of sequence files. The program Loadis runs on a Macintosh platform and parses the feature data extracted from the sequence files by Factura to the Unix based *Staphylococcus aureus* relational database. Assembly of contigs (and whole genome sequences) is accomplished by retrieving a specific set of sequence files and their associated features using extrseq, a Unix utility for retrieving sequences from an SQL database. The resulting sequence file is processed by seq\_filter to trim portions of the sequences with more than 2% ambiguous nucleotides. The sequence files were assembled using TIGR Assembler, an assembly engine designed at The Institute for Genomic Research (TIGR\*) for rapid and accurate assembly of thousands of sequence fragments. The collection of contigs generated by the assembly step is loaded into the database with the lassie program. Identification of open reading frames (ORFs) is accomplished by processing contigs with zorf. The ORFs are searched against *S. aureus* sequences from Genbank and against all protein sequences using the BLASTN and BLASTP programs, described in Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990)). Results of the ORF determination and similarity searching steps were loaded into the database. As described below, some results of the determination and the searches are set out in Tables 1-3.

The present invention is based on the sequencing of fragments of the *Staphylococcus aureus* genome and analysis of the sequences. The primary nucleotide sequences generated by sequencing the fragments are provided in SEQ ID NOS 1-5,191. (As used herein, the "primary sequence" refers to the nucleotide sequence represented by the IUPAC nomenclature system.)

In addition to the aforementioned *Staphylococcus aureus* polynucleotide and polynucleotide sequences, the present invention provides the nucleotide sequences of SEQ ID NOS 1-5,191, or representative fragments thereof, in a form which can be readily used, analyzed, and interpreted by a skilled artisan.

As used herein, a "representative fragment of the nucleotide sequence depicted in SEQ ID NOS 1-5,191" refers to any portion of the SEQ ID NOS 1-5,191 which is not presently represented within a publicly available database. Preferred representative fragments of the present invention are *Staphylococcus aureus* open reading frames (ORFs\*), expression modulating fragment (EMFs\*) and fragments which can be used to diagnose the presence of *Staphylococcus aureus* in sample ("DFs"). A non-limiting identification of preferred representative fragments is provided in Tables 1-3.

As discussed in detail below, the information provided in SEQ ID NOS 1-5,191 and in Tables 1-3 together with routine cloning, synthesis, sequencing and assay methods will enable those skilled in the art to clone and sequence all "representative fragments" of interest, including open reading frames encoding a large variety of *Staphylococcus aureus* proteins.

While the presently disclosed sequences of SEQ ID NOS 1-5,191 are highly accurate, sequencing techniques are not perfect and, in relatively rare instances, further investigation of a fragment or sequence of the invention may reveal a nucleotide sequence error present in a nucleotide sequence disclosed in SEQ ID NOS 1-5,191. However, once the present invention is made available (*i.e.*, once the information in SEQ ID NOS 1-5,191 and Tables 1-3 has been made available), resolving a rare sequencing error in SEQ ID NOS 1-5,191 will be well within the skill of the art. The present disclosure makes available sufficient sequence information to allow any of the described contigs or portions thereof to be obtained readily by straightforward application of routine techniques. Further sequencing of such polynucleotide may proceed in like manner using manual and automated sequencing methods which are employed ubiquitous in the art. Nucleotide sequence editing software is publicly available. For example, Applied Biosystem's (AB) AutoAssembler can be used as an aid during visual inspection of nucleotide sequences. By employing such routine techniques potential errors readily may be identified and the correct sequence then may be ascertained by targeting further sequencing effort, also of a routine nature, to the region containing the potential error.

Even if all of the very rare sequencing errors in SEQ ID NOS 1-5,191 were corrected, the resulting nucleotide sequences would still be at least 95% identical, nearly all would be at least 99% identical, and the great majority would be at least 99.9% identical to the nucleotide sequences of SEQ ID NOS 1-5,191.

As discussed elsewhere herein, polynucleotides of the present invention readily may be obtained by routine application of well known and standard procedures for cloning and sequencing DNA. Detailed methods for obtaining

libraries and for sequencing are provided below, for instance. A wide variety of *Staphylococcus aureus* strains that can be used to prepare *S aureus* genomic DNA for cloning and for obtaining polynucleotides of the present invention are available to the public from recognized depository institutions, such as the American Type Culture Collection (ATCC").

The nucleotide sequences of the genomes from different strains of *Staphylococcus aureus* differ somewhat. However, the nucleotide sequences of the genomes of all *Staphylococcus aureus* strains will be at least 95% identical, in corresponding part, to the nucleotide sequences provided in SEQ ID NOS:1-5,191. Nearly all will be at least 99% identical and the great majority will be 99.9% identical.

Thus, the present invention further provides nucleotide sequences which are at least 95%, preferably 99% and most preferably 99.9% identical to the nucleotide sequences of SEQ ID NOS 1-5,191, in a form which can be readily used, analyzed and interpreted by the skilled artisan.

Methods for determining whether a nucleotide sequence is at least 95%, at least 99% or at least 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-5,191 are routine and readily available to the skilled artisan. For example, the well known fasta algorithm described in Pearson and Lipman, *Proc Natl Acad Sci USA* 85:2444 (1988) can be used to generate the percent identity of nucleotide sequences. The BLASTN program also can be used to generate an identity score of polynucleotides compared to one another.

### COMPUTER RELATED EMBODIMENTS

The nucleotide sequences provided in SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a polynucleotide sequence of SEQ ID NOS 1-5,191 may be "provided" in a variety of mediums to facilitate use thereof. As used herein "provided" refers to a manufacture, other than an isolated nucleic acid molecule, which contains a nucleotide sequence of the present invention, i.e., a nucleotide sequence provided in SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a polynucleotide of SEQ ID NOS:1-5,191. Such a manufacture provides a large portion of the *Staphylococcus aureus* genome and parts thereof (e.g., a *Staphylococcus aureus* open reading frame (ORF)) in a form which allows a skilled artisan to examine the manufacture using means not directly applicable to examining the *Staphylococcus aureus* genome or a subset thereof as it exists in nature or in purified form.

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories, such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. Likewise, it will be clear to those of skill how additional computer readable media that may be developed also can be used to create analogous manufactures having recorded thereon a nucleotide sequence of the present invention.

As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially available software such as WordPerfect and MicroSoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data-processor structuring formats (e.g., text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. Thus, by providing in computer readable form the nucleotide sequences of SEQ ID NOS 1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a sequence of SEQ ID NOS:1-5,191 the present invention enables the skilled artisan routinely to access the provided sequence information for a wide variety of purposes.

The examples which follow demonstrate how software which implements the BLAST (Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990)) and BLAZE (Brutlag *et al.*, *Comp. Chem.* 17:203-207 (1993)) search algorithms on a Sybase system was used to identify open reading frames (ORFs) within the *Staphylococcus aureus* genome which contain homology to ORFs or proteins from both *Staphylococcus aureus* and from other organisms. Among the ORFs discussed

herein are protein encoding fragments of the *Staphylococcus aureus* genome useful in producing commercially important proteins, such as enzymes used in fermentation reactions and in the production of commercially useful metabolites.

The present invention further provides systems, particularly computer-based systems, which contain the sequence information described herein. Such systems are designed to identify, among other things, commercially important fragments of the *Staphylococcus aureus* genome.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based system are suitable for use in the present invention.

As stated above, the computer-based systems of the present invention comprise a data storage means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means.

As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of the present genomic sequences which match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBI). A skilled artisan can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems.

As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzymic active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present invention. A preferred format for an output means ranks fragments of the *Staphylococcus aureus* genomic sequences possessing varying degrees of homology to the target sequence or target motif. Such presentation provides a skilled artisan with a ranking of sequences which contain various amounts of the target sequence or target motif and identifies the degree of homology contained in the identified fragment.

A variety of comparing means can be used to compare a target sequence or target motif with the data storage means to identify sequence fragments of the *Staphylococcus aureus* genome. In the present examples, implementing software which implement the BLAST and BLAZE algorithms, described in Altschul *et al.*, *J. Mol. Biol.* **215**, 403-410 (1990), was used to identify open reading frames within the *Staphylococcus aureus* genome. A skilled artisan can readily recognize that any one of the publicly available homology search programs can be used as the search means for the computer-based systems of the present invention. Of course, suitable proprietary systems that may be known to those of skill also may be employed in this regard.

Figure 1 provides a block diagram of a computer system illustrative of embodiments of this aspect of present invention. The computer system 102 includes a processor 106 connected to a bus 104. Also connected to the bus 104 are a main memory 108 (preferably implemented as random access memory, RAM) and a variety of secondary storage devices 110, such as a hard drive 112 and a removable medium storage device 114. The removable medium storage device 114 may represent, for example, a floppy disk drive, a CD-ROM drive, a magnetic tape drive, etc. A removable storage medium 116 (such as a floppy disk, a compact disk, a magnetic tape, etc.) containing control logic and/or data recorded therein may be inserted into the removable medium storage device 114. The computer system 102 includes appropriate software for reading the control logic and/or the data from the removable medium storage device 114, once it is inserted into the removable medium storage device 114.

A nucleotide sequence of the present invention may be stored in a well known manner in the main memory 108, any of the secondary storage devices 110, and/or a removable storage medium 116. During execution, software for

accessing and processing the genomic sequence (such as search tools, comparing tools, etc.) reside in main memory 108, in accordance with the requirements and operating parameters of the operating system, the hardware system and the software program or programs.

## 5 BIOCHEMICAL EMBODIMENTS

Other embodiments of the present invention are directed to fragments of the *Staphylococcus aureus* genome, preferably to isolated fragments. The fragments of the *Staphylococcus aureus* genome of the present invention include, but are not limited to fragments which encode peptides, hereinafter open reading frames (ORFs), fragments which  
10 modulate the expression of an operably linked ORF, hereinafter expression modulating fragments (EMFs) and fragments which can be used to diagnose the presence of *Staphylococcus aureus* in a sample, hereinafter diagnostic fragments (DFs)

As used herein, an "isolated nucleic acid molecule" or an "isolated fragment of the *Staphylococcus aureus* genome" refers to a nucleic acid molecule possessing a specific nucleotide sequence which has been subjected to purification means to reduce, from the composition, the number of compounds which are normally associated with the composition. Particularly, the term refers to the nucleic acid molecules having the sequences set out in SEQ ID NOS 1-5,191, to representative fragments thereof as described above, to polynucleotides at least 95%, preferably at least 99% and especially preferably at least 99.9% identical in sequence thereto, also as set out above.

A variety of purification means can be used to generate the isolated fragments of the present invention. These  
20 include, but are not limited to methods which separate constituents of a solution based on charge, solubility, or size.

In one embodiment, *Staphylococcus aureus* DNA can be mechanically sheared to produce fragments of 15-20 kb in length. These fragments can then be used to generate an *Staphylococcus aureus* library by inserting them into lambda clones as described in the Examples below. Primers flanking, for example, an ORF, such as those enumerated in Tables 1-3 can then be generated using nucleotide sequence information provided in SEQ ID NOS: 1-5,191. Well  
25 known and routine techniques of PCR cloning then can be used to isolate the ORF from the lambda DNA library of *Staphylococcus aureus* genomic DNA. Thus, given the availability of SEQ ID NOS 1-5,191, the information in Tables 1, 2 and 3, and the information that may be obtained readily by analysis of the sequences of SEQ ID NOS 1-5,191 using methods set out above, those of skill will be enabled by the present disclosure to isolate any ORF-containing or other nucleic acid fragment of the present invention.

The isolated nucleic acid molecules of the present invention include, but are not limited to single stranded and double stranded DNA, and single stranded RNA.

As used herein, an "open reading frame," ORF, means a series of triplets coding for amino acids without any termination codons and is a sequence translatable into protein.

Tables 1, 2 and 3 list ORFs in the *Staphylococcus aureus* genomic contigs of the present invention that were  
35 identified as putative coding regions by the GeneMark software using organism-specific second-order Markov probability transition matrices. It will be appreciated that other criteria can be used, in accordance with well known analytical methods, such as those discussed herein, to generate more inclusive, more restrictive or more selective lists.

Table 1 sets out ORFs in the *Staphylococcus aureus* contigs of the present invention that are at least 80 amino acids long and over a continuous region of at least 50 bases which are 95% or more identical (by BLAST analysis) to an *S. aureus* nucleotide sequence available through Genbank in November 1996.

Table 2 sets out ORFs in the *Staphylococcus aureus* contigs of the present invention that are not in Table 1 and match, with a BLASTP probability score of 0.01 or less, a polypeptide sequence available through Genbank by September 1996.

Table 3 sets out ORFs in the *Staphylococcus aureus* contigs of the present invention that do not match significantly,  
45 by BLASTP analysis, a polypeptide sequence available through Genbank by September 1996.

In each table, the first and second columns identify the ORF by, respectively, contig number and ORF number within the contig; the third column indicates the reading frame, taking the first 5' nucleotide of the contig as the start of the +1 frame; the fourth column indicates the first nucleotide of the ORF, counting from the 5' end of the contig strand; and the fifth column indicates the length of each ORF in nucleotides.

In Tables 1 and 2, column six, lists the "Reference" for the closest matching sequence available through Genbank. These reference numbers are the databases entry numbers commonly used by those of skill in the art, who will be familiar with their denominators. Descriptions of the nomenclature are available from the National Center for Biotechnology Information. Column seven in Tables 1 and 2 provides the gene name\* of the matching sequence; column eight provides the BLAST identity\* score from the comparison of the ORF and the homologous gene; and column nine  
55 indicates the length in nucleotides of the highest scoring segment pair\* identified by the BLAST identity analysis.

In Table 3, the last column, column six, indicates the length of each ORF in amino acid residues.

The concepts of percent identity and percent similarity of two polypeptide sequences is well understood in the art. For example, two polypeptides 10 amino acids in length which differ at three amino acid positions (e.g., at positions

1, 3 and 5) are said to have a percent identity of 70%. However, the same two polypeptides would be deemed to have a percent similarity of 80% if, for example at position 5, the amino acids moieties, although not identical, were "similar" (i.e., possessed similar biochemical characteristics). Many programs for analysis of nucleotide or amino acid sequence similarity, such as fasta and BLAST specifically list per cent identity of a matching region as an output parameter. Thus, for instance, Tables 1 and 2 herein enumerate the per cent identity of the highest scoring segment pair in each ORF and its listed relative. Further details concerning the algorithms and criteria used for homology searches are provided below and are described in the pertinent literature highlighted by the citations provided below.

It will be appreciated that other criteria can be used to generate more inclusive and more exclusive listings of the types set out in the tables. As those of skill will appreciate, narrow and broad searches both are useful. Thus, a skilled artisan can readily identify ORFs in contigs of the *Staphylococcus aureus* genome other than those listed in Tables 1-3, such as ORFs which are overlapping or encoded by the opposite strand of an identified ORF in addition to those ascertainable using the computer-based systems of the present invention.

As used herein, an "expression modulating fragment," EMF, means a series of nucleotide molecules which modulates the expression of an operably linked ORF or EMF.

As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF. EMFs include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of EMFs are fragments which induce the expression of an operably linked ORF in response to a specific regulatory factor or physiological event.

EMF sequences can be identified within the contigs of the *Staphylococcus aureus* genome by their proximity to the ORFs provided in Tables 1-3. An intergenic segment, or a fragment of the intergenic segment, from about 10 to 200 nucleotides in length, taken from any one of the ORFs of Tables 1-3 will modulate the expression of an operably linked ORF in a fashion similar to that found with the naturally linked ORF sequence. As used herein, an "intergenic segment" refers to fragments of the *Staphylococcus aureus* genome which are between two ORF(s) herein described. EMFs also can be identified using known EMFs as a target sequence or target motif in the computer-based systems of the present invention. Further, the two methods can be combined and used together.

The presence and activity of an EMF can be confirmed using an EMF trap vector. An EMF trap vector contains a cloning site linked to a marker sequence. A marker sequence encodes an identifiable phenotype, such as antibiotic resistance or a complementing nutrition auxotrophic factor, which can be identified or assayed when the EMF trap vector is placed within an appropriate host under appropriate conditions. As described above, a EMF will modulate the expression of an operably linked marker sequence. A more detailed discussion of various marker sequences is provided below.

A sequence which is suspected as being an EMF is cloned in all three reading frames in one or more restriction sites upstream from the marker sequence in the EMF trap vector. The vector is then transformed into an appropriate host using known procedures and the phenotype of the transformed host is examined under appropriate conditions. As described above, an EMF will modulate the expression of an operably linked marker sequence.

As used herein, a "diagnostic fragment," DF, means a series of nucleotide molecules which selectively hybridize to *Staphylococcus aureus* sequences. DFs can be readily identified by identifying unique sequences within contigs of the *Staphylococcus aureus* genome, such as by using well-known computer analysis software, and by generating and testing probes or amplification primers consisting of the DF sequence in an appropriate diagnostic format which determines amplification or hybridization selectivity.

The sequences falling within the scope of the present invention are not limited to the specific sequences herein described, but also include allelic and species variations thereof. Allelic and species variations can be routinely determined by comparing the sequences provided in SEQ ID NOS: 1-5, 191, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably 99% and most preferably 99.9% identical to SEQ ID NOS: 1-5, 191, with a sequence from another isolate of the same species.

Furthermore, to accommodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the nucleic acid sequences mentioned above. In other words, in the coding region of an ORF, substitution of one codon for another which encodes the same amino acid is expressly contemplated.

Any specific sequence disclosed herein can be readily screened for errors by resequencing a particular fragment, such as an ORF, in both directions (i.e., sequence both strands). Alternatively, error screening can be performed by sequencing corresponding polynucleotides of *Staphylococcus aureus* origin isolated by using part or all of the fragments in question as a probe or primer.

Each of the ORFs of the *Staphylococcus aureus* genome disclosed in Tables 1, 2 and 3, and the EMFs found 5' to the ORFs, can be used as polynucleotide reagents in numerous ways. For example, the sequences can be used as diagnostic probes or diagnostic amplification primers to detect the presence of a specific microbe in a sample, particular *Staphylococcus aureus*. Especially preferred in this regard are ORF such as those of Table 3, which do not match previously characterized sequences from other organisms and thus are most likely to be highly selective for *Staphylococcus aureus*. Also particularly preferred are ORFs that can be used to distinguish between strains of *Sta-*



*phylococcus aureus*, particularly those that distinguish medically important strain, such as drug-resistant strains

In addition, the fragments of the present invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA. Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Information from the sequences of the present invention can be used to design antisense and triple helix-forming oligonucleotides. Polynucleotides suitable for use in these methods are usually 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription, for triple-helix formation or to the mRNA itself, for antisense inhibition. Both techniques have been demonstrated to be effective in model systems, and the requisite techniques are well known and involve routine procedures. Triple helix techniques are discussed in, for example, Lee *et al.*, *Nucl. Acids Res.* **6**: 3073 (1979); Cooney *et al.*, *Science* **241**: 456 (1988); and Dervan *et al.*, *Science* **251**: 1360 (1991). Antisense techniques in general are discussed in, for instance, Okano, *J. Neurochem.* **56**: 560 (1991) and OLIGODEOXYNUCLEOTIDES AS ANTISENSE INHIBITORS OF GENE EXPRESSION, CRC Press, Boca Raton, FL (1988).

The present invention further provides recombinant constructs comprising one or more fragments of the *Staphylococcus aureus* genomic fragments and contigs of the present invention. Certain preferred recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a fragment of the *Staphylococcus aureus* genome has been inserted, in a forward or reverse orientation. In the case of a vector comprising one of the ORFs of the present invention, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the ORF. For vectors comprising the EMFs of the present invention, the vector may further comprise a marker sequence or heterologous ORF operably linked to the EMF.

Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present invention. The following vectors are provided by way of example. Useful bacterial vectors include phagescript, PsiX174, pBluescript SK and KS (+ and -), pNH8a, pNH16a, pNH18a, pNH46a (available from Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (available from Pharmacia). Useful eukaryotic vectors include pWLneo, pSV2cat, pOG44, pXT1, pSG (available from Stratagene) pSVK3, pBPV, pMSG, pSVL (available from Pharmacia).

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein- I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

The present invention further provides host cells containing any one of the isolated fragments of the *Staphylococcus aureus* genomic fragments and contigs of the present invention, wherein the fragment has been introduced into the host cell using known methods. The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or a procaryotic cell, such as a bacterial cell.

A polynucleotide of the present invention, such as a recombinant construct comprising an ORF of the present invention, may be introduced into the host by a variety of well established techniques that are standard in the art, such as calcium phosphate transfection, DEAE dextran mediated transfection and electroporation, which are described in, for instance, Davis, L. *et al.*, BASIC METHODS IN MOLECULAR BIOLOGY (1986).

A host cell containing one of the fragments of the *Staphylococcus aureus* genomic fragments and contigs of the present invention, can be used in conventional manners to produce the gene product encoded by the isolated fragment (in the case of an ORF) or can be used to produce a heterologous protein under the control of the EMF.

The present invention further provides isolated polypeptides encoded by the nucleic acid fragments of the present invention or by degenerate variants of the nucleic acid fragments of the present invention. By "degenerate variant" is intended nucleotide fragments which differ from a nucleic acid fragment of the present invention (e.g., an ORF) by nucleotide sequence but, due to the degeneracy of the Genetic Code, encode an identical polypeptide sequence.

Preferred nucleic acid fragments of the present invention are the ORFs depicted in Tables 2 and 3 which encode proteins.

A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. This is particularly useful in producing small peptides and fragments of larger polypeptides. Such short fragments as may be obtained most readily by synthesis are useful, for example, in generating antibodies against the native polypeptide, as discussed further below.

In an alternative method, the polypeptide or protein is purified from bacterial cells which naturally produce the polypeptide or protein. One skilled in the art can readily employ well-known methods for isolating polypeptides and proteins to isolate and purify polypeptides or proteins of the present invention produced naturally by a bacterial strain, or by other methods. Methods for isolation and purification that can be employed in this regard include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography, ion-exchange chromatography, and immu-

no-affinity chromatography

The polypeptides and proteins of the present invention also can be purified from cells which have been altered to express the desired polypeptide or protein. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally does not produce or which the cell normally produces at a lower level. Those skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention

Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells, CV-1 cell, COS cells, and Sf9 cells, as well as prokaryotic host such as *E. coli* and *B. subtilis*. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level

"Recombinant," as used herein, means that a polypeptide or protein is derived from recombinant (e.g., microbial or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (e.g., yeast) expression systems. As a product, "recombinant microbial" defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Polypeptides or proteins expressed in most bacterial cultures, e.g., *E. coli*, will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern different from that expressed in mammalian cells

"Nucleotide sequence" refers to a heteropolymer of deoxyribonucleotides. Generally, DNA segments encoding the polypeptides and proteins provided by this invention are assembled from fragments of the *Staphylococcus aureus* genome and short oligonucleotide linkers, or from a series of oligonucleotides, to provide a synthetic gene which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon.

"Recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. The expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic regulatory elements necessary for gene expression in the host, including elements required to initiate and maintain transcription at a level sufficient for suitable expression of the desired polypeptide, including, for example, promoters and, where necessary, an enhancers and a polyadenylation signal; (2) a structural or coding sequence which is transcribed into mRNA and translated into protein; and (3) appropriate signals to initiate translation at the beginning of the desired coding region and terminate translation at its end. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an N-terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

"Recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extra chromosomally. The cells can be prokaryotic or eukaryotic. Recombinant expression systems as defined herein will express heterologous polypeptides or proteins upon induction of the regulatory elements linked to the DNA segment or synthetic gene to be expressed.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described in Sambrook *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL, 2<sup>nd</sup> Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference in its entirety.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), alpha-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product.

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and, when desirable, provide amplification within the host.

Suitable prokaryotic hosts for transformation include strains of *Staphylococcus aureus*, *E. coli*, *B. subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*. Others

may also be employed as a matter of choice

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (available from Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (available from Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter, where it is inducible, is derepressed or induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period to provide for expression of the induced gene product. Thereafter cells are typically harvested, generally by centrifugation, disrupted to release expressed protein, generally by physical or chemical means, and the resulting crude extract is retained for further purification.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described in Gluzman, *Cell* 23: 175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines.

Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

Recombinant polypeptides and proteins produced in bacterial culture is usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

An additional aspect of the invention includes *Staphylococcus aureus* polypeptides which are useful as immunodiagnostic antigens and/or immunoprotective vaccines, collectively "immunologically useful polypeptides". Such immunologically useful polypeptides may be selected from the ORFs disclosed herein based on techniques well known in the art and described elsewhere herein. The inventors have used the following criteria to select several immunologically useful polypeptides:

As is known in the art, an amino terminal type I signal sequence directs a nascent protein across the plasma and outer membranes to the exterior of the bacterial cell. Such outer membrane polypeptides are expected to be immunologically useful. According to Izard, J. W. et al., *Mol. Microbiol.* 13: 765-773; (1994), polypeptides containing type I signal sequences contain the following physical attributes: The length of the type I signal sequence is approximately 15 to 25 primarily hydrophobic amino acid residues with a net positive charge in the extreme amino terminus; the central region of the signal sequence must adopt an alpha-helical conformation in a hydrophobic environment; and the region surrounding the actual site of cleavage is ideally six residues long, with small side-chain amino acids in the -1 and -3 positions.

Also known in the art is the type IV signal sequence which is an example of the several types of functional signal sequences which exist in addition to the type I signal sequence detailed above. Although functionally related, the type IV signal sequence possesses a unique set of biochemical and physical attributes (Strom, M. S. and Lory, S., *J. Bacteriol.* 174: 7345-7351; 1992). These are typically six to eight amino acids with a net basic charge followed by an additional sixteen to thirty primarily hydrophobic residues. The cleavage site of a type IV signal sequence is typically after the initial six to eight amino acids at the extreme amino terminus. In addition, all type IV signal sequences contain a phenylalanine residue at the +1 site relative to the cleavage site.

Studies of the cleavage sites of twenty-six bacterial lipoprotein precursors has allowed the definition of a consensus amino acid sequence for lipoprotein cleavage. Nearly three-fourths of the bacterial lipoprotein precursors examined contained the sequence L-(A/S)-(G/A)-C at positions -3 to +1, relative to the point of cleavage (Hayashi, S. and Wu, H. C. *Lipoproteins in bacteria*, J. Bioenerg. Biomembr. 22: 451-471; 1990).

It is well known that most anchored proteins found on the surface of gram-positive bacteria possess a highly conserved carboxy terminal sequence. More than fifty such proteins from organisms such as *S. pyogenes*, *S. mutans*, *E. faecalis*, *S. pneumoniae*, and others, have been identified based on their extracellular location and carboxy terminal amino acid sequence (Fischetti, V. A. Gram-positive commensal bacteria deliver antigens to elicit mucosal and systemic immunity. *ASM News* 62: 405410; 1996). The conserved region is comprised of six charged amino acids at the extreme carboxy terminus coupled to 15-20 hydrophobic amino acids presumed to function as a transmembrane domain. Immediately adjacent to the transmembrane domain is a six amino acid sequence conserved in nearly all proteins ex-

aminated. The amino acid sequence of this region is L-P-X-T-G-X, where X is any amino acid

Amino acid sequence similarities to proteins of known function by BLAST enables the assignment of putative functions to novel amino acid sequences and allows for the selection of proteins thought to function outside the cell wall. Such proteins are well known in the art and include "lipoprotein", "periplasmic", or "antigen"

5 An algorithm for selecting antigenic and immunogenic *Staphylococcus aureus* polypeptides including the foregoing criteria was developed by the present inventors. Use of the algorithm by the inventors to select immunologically useful *Staphylococcus aureus* polypeptides resulted in the selection of several ORF's which are predicted to be outer-membrane-associated proteins. These proteins are identified in Table 4, below, and shown in the Sequence Listing as SEQ ID NOS 5, 192 to 5, 255. Thus the amino acid sequence of each of several antigenic *Staphylococcus aureus* polypeptides  
10 listed in Table 4 can be determined, for example, by locating the amino acid sequence of the ORF in the Sequence Listing. Likewise the polynucleotide sequence encoding each ORF can be found by locating the corresponding polynucleotide SEQ ID in Tables 1, 2, or 3, and finding the corresponding nucleotide sequence in the sequence listing.

As will be appreciated by those of ordinary skill in the art, although a polypeptide representing an entire ORF may be the closest approximation to a protein found *in vivo*, it is not always technically practical to express a complete ORF  
15 *in vitro*. It may be very challenging to express and purify a highly hydrophobic protein by common laboratory methods. As a result, the immunologically useful polypeptides described herein as SEQ ID NOS 5, 192-5, 255 may have been modified slightly to simplify the production of recombinant protein, and are the preferred embodiments. In general, nucleotide sequences which encode highly hydrophobic domains, such as those found at the amino terminal signal sequence, are excluded for enhanced *in vitro* expression of the polypeptides. Furthermore, any highly hydrophobic  
20 amino acid sequences occurring at the carboxy terminus are also excluded. Such truncated polypeptides include for example the mature forms of the polypeptides expected to exist in nature.

Those of ordinary skill in the art can identify soluble portions the polypeptide identified in Table 4, and in the case of truncated polypeptides sequences shown as SEQ ID NOS 5, 192-5, 255, may obtain the complete predicted amino acid sequence of each polypeptide by translating the corresponding polynucleotides sequences of the corresponding  
25 ORF listed in Tables 1, 2 and 3 and found in the sequence listing.

Accordingly, polypeptides comprising the complete amino acid of an immunologically useful polypeptide selected from the group of polypeptides encoded by the ORF's identified in Table 4, or an amino acid sequence at least 95% identical thereto, preferably at least 97% identical thereto, and most preferably at least 99% identical thereto form an embodiment of the invention; in addition polypeptides comprising an amino acid sequence selected from the group of  
30 amino acid sequences shown in the sequence listing as SEQ ID NOS 5, 191-5, 255, or an amino acid sequence at least 95% identical thereto, preferably at least 97% identical thereto and most preferably at least 99% identical thereto, form an embodiment of the invention. Polynucleotides encoding the foregoing polypeptides also form part of the present invention.

In another aspect, the invention provides a peptide or polypeptide comprising an epitope-bearing portion of a polypeptide of the invention, particularly those epitope-bearing portions (antigenic regions) identified in Table 4. The epitope-bearing portion is an immunogenic or antigenic epitope of a polypeptide of the invention. An "immunogenic epitope" is defined as a part of a protein that elicits an antibody response when the whole protein is the immunogen. On the other hand, a region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." The number of immunogenic epitopes of a protein generally is less than the number of antigenic epitopes. See, for  
40 instance, Geysen et al., Proc. Natl. Acad. Sci. USA 81:3998-4002 (1983).

As to the selection of peptides or polypeptides bearing an antigenic epitope (i.e., that contain a region of a protein molecule to which an antibody can bind), it is well known in that art that relatively short synthetic peptides that mimic part of a protein sequence are routinely capable of eliciting an antiserum that reacts with the partially mimicked protein. See, for instance, Sutcliffe, J. G., Shinnick, T. M., Green, N. and Learner, R. A. (1983) "Antibodies that react with predetermined sites on proteins", Science, 219:660-666. Peptides capable of eliciting protein-reactive sera are frequently represented in the primary sequence of a protein, can be characterized by a set of simple chemical rules, and are confined neither to immunodominant regions of intact proteins (i.e., immunogenic epitopes) nor to the amino or carboxyl terminals. Antigenic epitope-bearing peptides and polypeptides of the invention are therefore useful to raise antibodies, including monoclonal antibodies, that bind specifically to a polypeptide of the invention. See, for instance,  
50 Wilson et al., Cell 37:767-778 (1984) at 777.

Antigenic epitope-bearing peptides and polypeptides of the invention preferably contain a sequence of at least seven, more preferably at least nine and most preferably between about 15 to about 30 amino acids contained within the amino acid sequence of a polypeptide of the invention. Non-limiting examples of antigenic polypeptides or peptides that can be used to generate *S. aureus* specific antibodies include: a polypeptide comprising peptides shown in Table  
55 4 below. These polypeptide fragments have been determined to bear antigenic epitopes of indicated *S. aureus* proteins by the analysis of the Jameson-Wolf antigenic index, a representative sample of which is shown in Figure 3.

The epitope-bearing peptides and polypeptides of the invention may be produced by any conventional means. See, e.g., Houghten, R. A. (1985) General method for the rapid solid-phase synthesis of large numbers of peptides;

specificity of antigen-antibody interaction at the level of individual amino acids. Proc. Natl. Acad. Sci. USA 82: 5131-5135. This "Simultaneous Multiple Peptide Synthesis (SMPS)" process is further described in U.S. Patent No. 4,631,211 to Houghten et al. (1986). Epitope-bearing peptides and polypeptides of the invention are used to induce antibodies according to methods well known in the art. See, for instance, Sutcliffe et al., supra; Wilson et al., supra; Chow, M. et al., Proc. Natl. Acad. Sci. USA 82:910-914, and Bittle, F. J. et al., J. Gen. Virol. 66:2347-2354 (1985).

Immunogenic epitope-bearing peptides of the invention, i.e., those parts of a protein that elicit an antibody response when the whole protein is the immunogen, are identified according to methods known in the art. See, for instance, Geysen et al. supra. Further still, U.S. Patent No. 5,194,392 to Geysen (1990) describes a general method of detecting or determining the sequence of monomers (amino acids or other compounds) which is a topological equivalent of the epitope (i.e., a "mimotope") which is complementary to a particular paratope (antigen binding site) of an antibody of interest. More generally, U.S. Patent No. 4,433,092 to Geysen (1989) describes a method of detecting or determining a sequence of monomers which is a topographical equivalent of a ligand which is complementary to the ligand binding site of a particular receptor of interest. Similarly, U.S. Patent No. 5,480,971 to Houghten, R. A. et al. (1996) on Per-alkylated Oligopeptide Mixtures discloses linear C1-C7-alkyl peralkylated oligopeptides and sets and libraries of such peptides, as well as methods for using such oligopeptide sets and libraries for determining the sequence of a per-alkylated oligopeptide that preferentially binds to an acceptor molecule of interest. Thus, non-peptide analogs of the epitope-bearing peptides of the invention also can be made routinely by these methods.

Table 4 lists immunologically useful polypeptides identified by an algorithm which locates novel *Staphylococcus aureus* outer membrane proteins, as is described above. Also listed are epitopes or "antigenic regions" of each of the identified polypeptides. The antigenic regions, or epitopes, are delineated by two numbers x-y, where x is the number of the first amino acid in the open reading frame included within the epitope and y is the number of the last amino acid in the open reading frame included within the epitope. For example, the first epitope in ORF 168-6 is comprised of amino acids 36 to 45 of SEQ ID NO.5,192, as is described in Table 4. The inventors have identified several epitopes for each of the antigenic polypeptides identified in Table 4. Accordingly, forming part of the present invention are polypeptides comprising an amino acid sequence of one or more antigenic regions identified in Table 4. The invention further provides polynucleotides encoding such polypeptides.

The present invention further includes isolated polypeptides, proteins and nucleic acid molecules which are substantially equivalent to those herein described. As used herein, substantially equivalent can refer both to nucleic acid and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity between reference and subject sequences. For purposes of the present invention, sequences having equivalent biological activity and equivalent expression characteristics are considered substantially equivalent. For purposes of determining equivalence, truncation of the mature sequence should be disregarded.

The invention further provides methods of obtaining homologs from other strains of *Staphylococcus aureus*, of the fragments of the *Staphylococcus aureus* genome of the present invention and homologs of the proteins encoded by the ORFs of the present invention. As used herein, a sequence or protein of *Staphylococcus aureus* is defined as a homolog of a fragment of the *Staphylococcus aureus* fragments or contigs or a protein encoded by one of the ORFs of the present invention, if it shares significant homology to one of the fragments of the *Staphylococcus aureus* genome of the present invention or a protein encoded by one of the ORFs of the present invention. Specifically, by using the sequence disclosed herein as a probe or as primers, and techniques such as PCR cloning and colony/plaque hybridization, one skilled in the art can obtain homologs.

As used herein, two nucleic acid molecules or proteins are said to "share significant homology" if the two contain regions which possess greater than 85% sequence (amino acid or nucleic acid) homology. Preferred homologs in this regard are those with more than 90% homology. Especially preferred are those with 93% or more homology. Among especially preferred homologs those with 95% or more homology are particularly preferred. Very particularly preferred among these are those with 97% and even more particularly preferred among these are homologs with 99% or more homology. The most preferred homologs among these are those with 99.9% homology or more. It will be understood that, among measures of homology, identity is particularly preferred in this regard.

Region specific primers or probes derived from the nucleotide sequence provided in SEQ ID NOS 1-5,191 or from a nucleotide sequence at least 95%, particularly at least 99%, especially at least 99.5% identical to a sequence of SEQ ID NOS 1-5,191 can be used to prime DNA synthesis and PCR amplification, as well as to identify colonies containing cloned DNA encoding a homolog. Methods suitable to this aspect of the present invention are well known and have been described in great detail in many publications such as, for example, Innis et al., PCR PROTOCOLS, Academic Press, San Diego, CA (1990)).

When using primers derived from SEQ ID NOS 1-5,191 or from a nucleotide sequence having an aforementioned identity to a sequence of SEQ ID NOS 1-5,191, one skilled in the art will recognize that by employing high stringency conditions (e.g., annealing at 50-60°C in 6X SSPE and 50% formamide, and washing at 50-65°C in 0.5X SSPE) only sequences which are greater than 75% homologous to the primer will be amplified. By employing lower stringency

conditions (e.g., hybridizing at 35-37°C in 5X SSPC and 40-45% formamide, and washing at 42°C in 0.5X SSPC), sequences which are greater than 40-50% homologous to the primer will also be amplified.

When using DNA probes derived from SEQ ID NOS 1-5, 191, or from a nucleotide sequence having an aforementioned identity to a sequence of SEQ ID NOS 1-5, 191, for colony/plaque hybridization, one skilled in the art will recognize that by employing high stringency conditions (e.g., hybridizing at 50-65°C in 5X SSPC and 50% formamide, and washing at 50-65°C in 0.5X SSPC), sequences having regions which are greater than 90% homologous to the probe can be obtained, and that by employing lower stringency conditions (e.g., hybridizing at 35-37°C in 5X SSPC and 40-45% formamide, and washing at 42°C in 0.5X SSPC), sequences having regions which are greater than 35-45% homologous to the probe will be obtained.

Any organism can be used as the source for homologs of the present invention so long as the organism naturally expresses such a protein or contains genes encoding the same. The most preferred organism for isolating homologs are bacteria which are closely related to *Staphylococcus aureus*.

## ILLUSTRATIVE USES OF COMPOSITIONS OF THE INVENTION

Each ORF provided in Tables 1 and 2 is identified with a function by homology to a known gene or polypeptide. As a result, one skilled in the art can use the polypeptides of the present invention for commercial, therapeutic and industrial purposes consistent with the type of putative identification of the polypeptide. Such identifications permit one skilled in the art to use the *Staphylococcus aureus* ORFs in a manner similar to the known type of sequences for which the identification is made; for example, to ferment a particular sugar source or to produce a particular metabolite. A variety of reviews illustrative of this aspect of the invention are available, including the following reviews on the industrial use of enzymes, for example, BIOCHEMICAL ENGINEERING AND BIOTECHNOLOGY HANDBOOK, 2nd Ed., Macmillan Publications, Ltd. NY (1991) and BIOCATALYSTS IN ORGANIC SYNTHESIS, Tramper *et al.*, Eds., Elsevier Science Publishers, Amsterdam, The Netherlands (1985). A variety of exemplary uses that illustrate this and similar aspects of the present invention are discussed below.

### 1. Biosynthetic Enzymes

Open reading frames encoding proteins involved in mediating the catalytic reactions involved in intermediary and macromolecular metabolism, the biosynthesis of small molecules, cellular processes and other functions includes enzymes involved in the degradation of the intermediary products of metabolism, enzymes involved in central intermediary metabolism, enzymes involved in respiration, both aerobic and anaerobic, enzymes involved in fermentation, enzymes involved in ATP proton motor force conversion, enzymes involved in broad regulatory function, enzymes involved in amino acid synthesis, enzymes involved in nucleotide synthesis, enzymes involved in cofactor and vitamin synthesis, can be used for industrial biosynthesis.

The various metabolic pathways present in *Staphylococcus aureus* can be identified based on absolute nutritional requirements as well as by examining the various enzymes identified in Table 1-3 and SEQ ID NOS 1-5, 191.

Of particular interest are polypeptides involved in the degradation of intermediary metabolites as well as non-macromolecular metabolism. Such enzymes include amylases, glucose oxidases, and catalase.

Proteolytic enzymes are another class of commercially important enzymes. Proteolytic enzymes find use in a number of industrial processes including the processing of flax and other vegetable fibers, in the extraction, clarification and depectinization of fruit juices, in the extraction of vegetables' oil and in the maceration of fruits and vegetables to give unicellular fruits. A detailed review of the proteolytic enzymes used in the food industry is provided in Rombouts *et al.*, *Symbiosis* 21: 79 (1986) and Voragen *et al.* in BIOCATALYSTS IN AGRICULTURAL BIOTECHNOLOGY, Whitaker *et al.*, Eds., American Chemical Society Symposium Series 389, 93 (1989).

The metabolism of sugars is an important aspect of the primary metabolism of *Staphylococcus aureus*. Enzymes involved in the degradation of sugars, such as, particularly, glucose, galactose, fructose and xylose, can be used in industrial fermentation. Some of the important sugar transforming enzymes, from a commercial viewpoint, include sugar isomerases such as glucose isomerase. Other metabolic enzymes have found commercial use such as glucose oxidases which produces ketogulononic acid (KGA). KGA is an intermediate in the commercial production of ascorbic acid using the Reichstein's procedure, as described in Krueger *et al.*, *Biotechnology* 6(A), Rhine *et al.*, Eds., Verlag Press, Weinheim, Germany (1984).

Glucose oxidase (GOD) is commercially available and has been used in purified form as well as in an immobilized form for the deoxygenation of beer. See, for instance, Hartmeir *et al.*, *Biotechnology Letters* 1: 21 (1979). The most important application of GOD is the industrial scale fermentation of gluconic acid. Market for gluconic acids which are used in the detergent, textile, leather, photographic, pharmaceutical, food, feed and concrete industry, as described, for example, in Bigelis *et al.*, beginning on page 357 in GENE MANIPULATIONS AND FUNGI, Benett *et al.*, Eds., Academic Press, New York (1985). In addition to industrial applications, GOD has found applications in medicine for

quantitative determination of glucose in body fluids recently in biotechnology for analyzing syrups from starch and cellulose hydrosylates. This application is described in Owusu *et al.*, *Biochem. et Biophysica. Acta* 872, 83 (1986) for instance.

The main sweetener used in the world today is sugar which comes from sugar beets and sugar cane. In the field of industrial enzymes, the glucose isomerase process shows the largest expansion in the market today. Initially, soluble enzymes were used and later immobilized enzymes were developed (Krueger *et al.*, *Biotechnology. The Textbook of Industrial Microbiology*, Sinauer Associated Incorporated, Sunderland, Massachusetts (1990)). Today, the use of glucose-produced high fructose syrups is by far the largest industrial business using immobilized enzymes. A review of the industrial use of these enzymes is provided by Jorgensen, *Starch* 40 307 (1988).

Proteinases, such as alkaline serine proteinases, are used as detergent additives and thus represent one of the largest volumes of microbial enzymes used in the industrial sector. Because of their industrial importance, there is a large body of published and unpublished information regarding the use of these enzymes in industrial processes. (See Faultman *et al.*, *Acid Proteases Structure Function and Biology*, Tang, J., ed. Plenum Press, New York (1977) and Godfrey *et al.*, *Industrial Enzymes*, MacMillan Publishers, Surrey, UK (1983) and Hepner *et al.*, *Report Industrial Enzymes by 1990*, Hel Hepner & Associates, London (1986)).

Another class of commercially usable proteins of the present invention are the microbial lipases, described by for instance, Macrae *et al.*, *Philosophical Transactions of the Chiral Society of London* 310 227 (1985) and Poserke, *Journal of the American Oil Chemist Society* 61:1758 (1984). A major use of lipases is in the fat and oil industry for the production of neutral glycerides using lipase catalyzed inter-esterification of readily available triglycerides. Application of lipases include the use as a detergent additive to facilitate the removal of fats from fabrics in the course of the washing procedures.

The use of enzymes, and in particular microbial enzymes, as catalyst for key steps in the synthesis of complex organic molecules is gaining popularity at a great rate. One area of great interest is the preparation of chiral intermediates. Preparation of chiral intermediates is of interest to a wide range of synthetic chemists particularly those scientists involved with the preparation of new pharmaceuticals, agrochemicals, fragrances and flavors. (See Davies *et al.*, *Recent Advances in the Generation of Chiral Intermediates Using Enzymes*, CRC Press, Boca Raton, Florida (1990)). The following reactions catalyzed by enzymes are of interest to organic chemists: hydrolysis of carboxylic acid esters, phosphate esters, amides and nitriles, esterification reactions, trans-esterification reactions, synthesis of amides, reduction of alkanones and oxoalkanates, oxidation of alcohols to carbonyl compounds, oxidation of sulfides to sulfoxides, and carbon bond forming reactions such as the aldol reaction.

When considering the use of an enzyme encoded by one of the ORFs of the present invention for biotransformation and organic synthesis it is sometimes necessary to consider the respective advantages and disadvantages of using a microorganism as opposed to an isolated enzyme. Pros and cons of using a whole cell system on the one hand or an isolated partially purified enzyme on the other hand, has been described in detail by Bud *et al.*, *Chemistry in Britain* (1987), p. 127.

Amino transferases, enzymes involved in the biosynthesis and metabolism of amino acids, are useful in the catalytic production of amino acids. The advantages of using microbial based enzyme systems is that the amino transferase enzymes catalyze the stereo-selective synthesis of only L-amino acids and generally possess uniformly high catalytic rates. A description of the use of amino transferases for amino acid production is provided by Roselle-David, *Methods of Enzymology* 136 479 (1987).

Another category of useful proteins encoded by the ORFs of the present invention include enzymes involved in nucleic acid synthesis, repair, and recombination. A variety of commercially important enzymes have previously been isolated from members of *Staphylococcus aureus*. These include Sau3A and Sau96I.

## 2. Generation of Antibodies

As described here, the proteins of the present invention, as well as homologs thereof, can be used in a variety of procedures and methods known in the art which are currently applied to other proteins. The proteins of the present invention can further be used to generate an antibody which selectively binds the protein. Such antibodies can be either monoclonal or polyclonal antibodies, as well fragments of these antibodies, and humanized forms.

The invention further provides antibodies which selectively bind to one of the proteins of the present invention and hybridomas which produce these antibodies. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

In general, techniques for preparing polyclonal and monoclonal antibodies as well as hybridomas capable of producing the desired antibody are well known in the art (Campbell, A. M., *MONOCLONAL ANTIBODY TECHNOLOGY: LABORATORY TECHNIQUES IN BIOCHEMISTRY AND MOLECULAR BIOLOGY*, Elsevier Science Publishers, Amsterdam, The Netherlands (1984); St. Groth *et al.*, *J. Immunol. Methods* 35 1-21 (1980); Kohler and Milstein, *Nature* 256 495-497 (1975)); the trioma technique, the human B- cell hybridoma technique (Kozbor *et al.*, *Immunology Today*

4-72 (1983), pgs. 77-96 of Cole *et al.*, in MONOCLONAL ANTIBODIES AND CANCER THERAPY. Alan R. Liss, Inc. (1985)).

Any animal (mouse, rabbit, *etc.*) which is known to produce antibodies can be immunized with the pseudogene polypeptide. Methods for immunization are well known in the art. Such methods include subcutaneous or interperitoneal injection of the polypeptide. One skilled in the art will recognize that the amount of the protein encoded by the ORF of the present invention used for immunization will vary based on the animal which is immunized, the antigenicity of the peptide and the site of injection.

The protein which is used as an immunogen may be modified or administered in an adjuvant in order to increase the protein's antigenicity. Methods of increasing the antigenicity of a protein are well known in the art and include, but are not limited to coupling the antigen with a heterologous protein (such as globulin or galactosidase) or through the inclusion of an adjuvant during immunization.

For monoclonal antibodies, spleen cells from the immunized animals are removed, fused with myeloma cells, such as SP2/0-Ag14 myeloma cells, and allowed to become monoclonal antibody producing hybridoma cells.

Any one of a number of methods well known in the art can be used to identify the hybridoma cell which produces an antibody with the desired characteristics. These include screening the hybridomas with an ELISA assay, western blot analysis, or radioimmunoassay (Lutz *et al.*, *Exp. Cell Res.* 175: 109-124 (1988)).

Hybridomas secreting the desired antibodies are cloned and the class and subclass is determined using procedures known in the art (Campbell, A. M., *Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1984)).

Techniques described for the production of single chain antibodies (U. S. Patent 4,946,778) can be adapted to produce single chain antibodies to proteins of the present invention.

For polyclonal antibodies, antibody containing antisera is isolated from the immunized animal and is screened for the presence of antibodies with the desired specificity using one of the above-described procedures.

The present invention further provides the above-described antibodies in detectably labelled form. Antibodies can be detectably labelled through the use of radioisotopes, affinity labels (such as biotin, avidin, *etc.*), enzymatic labels (such as horseradish peroxidase, alkaline phosphatase, *etc.*) fluorescent labels (such as FITC or rhodamine, *etc.*), paramagnetic atoms, *etc.* Procedures for accomplishing such labelling are well-known in the art, for example see Sternberger *et al.*, *J. Histochem. Cytochem.* 18:315 (1970); Bayer, E. A. *et al.*, *Meth. Enzym.* 62:308 (1979); Engval, E. *et al.*, *Immunol.* 109:129 (1972); Goding, J. W. *J. Immunol. Meth.* 13:215 (1976).

The labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays to identify cells or tissues in which a fragment of the *Staphylococcus aureus* genome is expressed.

The present invention further provides the above-described antibodies immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, acrylic resins and such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known in the art (Weir, D. M. *et al.*, "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby, W. D. *et al.*, *Meth. Enzym.* 34 Academic Press, N. Y. (1974)). The immobilized antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays as well as for immunoaffinity purification of the proteins of the present invention.

### 3. Diagnostic Assays and Kits

The present invention further provides methods to identify the expression of one of the ORFs of the present invention, or homolog thereof, in a test sample, using one of the DFs, antigens or antibodies of the present invention.

In detail, such methods comprise incubating a test sample with one or more of the antibodies, or one or more of the DFs, or one or more antigens of the present invention and assaying for binding of the DFs, antigens or antibodies to components within the test sample.

Conditions for incubating a DF, antigen or antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the DF or antibody used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or immunological assay formats can readily be adapted to employ the DFs, antigens or antibodies of the present invention. Examples of such assays can be found in Chard, T., *An Introduction to Radioimmunoassay and Related Techniques*, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. *et al.*, *Techniques in Immunocytochemistry*, Academic Press, Orlando, FL Vol. 1 (1982) Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry*, PCT publication WO95/32291, and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1985), all of which are hereby incorporated herein by reference.

The test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as sputum, blood, serum, plasma, or urine. The test sample used in the above-described method will vary based



on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises (a) a first container comprising one of the Dfs, antigens or antibodies of the present invention; and (b) one or more other containers comprising one or more of the following wash reagents: reagents capable of detecting presence of a bound DF, antigen or antibody.

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the antibodies used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound antibody, antigen or DF.

Types of detection reagents include labelled nucleic acid probes, labelled secondary antibodies, or in the alternative, if the primary antibody is labelled, the enzymatic, or antibody binding reagents which are capable of reacting with the labelled antibody. One skilled in the art will readily recognize that the disclosed Dfs, antigens and antibodies of the present invention can be readily incorporated into one of the established kit formats which are well known in the art.

#### 4. Screening Assay for Binding Agents

Using the isolated proteins of the present invention, the present invention further provides methods of obtaining and identifying agents which bind to a protein encoded by one of the ORFs of the present invention or to one of the fragments and the *Staphylococcus aureus* fragment and contigs herein described.

In general, such methods comprise steps of

- (a) contacting an agent with an isolated protein encoded by one of the ORFs of the present invention, or an isolated fragment of the *Staphylococcus aureus* genome; and
- (b) determining whether the agent binds to said protein or said fragment

The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling techniques.

For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention.

Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed" when the agent is chosen based on the configuration of the particular protein. For example, one skilled in the art can readily adapt currently available procedures to generate peptides, pharmaceutical agents and the like capable of binding to a specific peptide sequence in order to generate rationally designed antipeptide peptides, for example see Hurby *et al.*, Application of Synthetic Peptides: Antisense Peptides," In Synthetic Peptides, A User's Guide, W. H. Freeman, NY (1992), pp. 289-307, and Kaspczak *et al.*, Biochemistry 28: 9230-8 (1989), or pharmaceutical agents, or the like.

In addition to the foregoing, one class of agents of the present invention, as broadly described, can be used to control gene expression through binding to one of the ORFs or EMFs of the present invention. As described above, such agents can be randomly screened or rationally designed/selected. Targeting the ORF or EMF allows a skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or multiple ORFs which rely on the same EMF for expression control.

One class of DNA binding agents are agents which contain base residues which hybridize or form a triple helix by binding to DNA or RNA. Such agents can be based on the classic phosphodiester ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivatives which have base attachment capacity.

Agents suitable for use in these methods usually contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee *et al.*, Nucl. Acids Res. 6: 3073 (1979); Cooney *et al.*, Science 241: 456 (1988), and Dervan *et al.*, Science 251: 1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56: 560 (1991). Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated

to be effective in model systems. Information contained in the sequences of the present invention can be used to design antisense and triple helix-forming oligonucleotides, and other DNA binding agents.

## 5. Pharmaceutical Compositions and Vaccines

The present invention further provides pharmaceutical agents which can be used to modulate the growth or pathogenicity of *Staphylococcus aureus*, or another related organism, *in vivo* or *in vitro*. As used herein, a "pharmaceutical agent" is defined as a composition of matter which can be formulated using known techniques to provide a pharmaceutical compositions. As used herein, the "pharmaceutical agents of the present invention" refers the pharmaceutical agents which are derived from the proteins encoded by the ORFs of the present invention or are agents which are identified using the herein described assays.

As used herein, a pharmaceutical agent is said to "modulate the growth or pathogenicity of *Staphylococcus aureus* or a related organism, *in vivo* or *in vitro*," when the agent reduces the rate of growth, rate of division, or viability of the organism in question. The pharmaceutical agents of the present invention can modulate the growth or pathogenicity of an organism in many fashions, although an understanding of the underlying mechanism of action is not needed to practice the use of the pharmaceutical agents of the present invention. Some agents will modulate the growth or pathogenicity by binding to an important protein thus blocking the biological activity of the protein, while other agents may bind to a component of the outer surface of the organism blocking attachment or rendering the organism more prone to act the bodies nature immune system. Alternatively, the agent may comprise a protein encoded by one of the ORFs of the present invention and serve as a vaccine. The development and use of vaccines derived from membrane associated polypeptides are well known in the art. The inventors have identified particularly preferred immunogenic *Staphylococcus aureus* polypeptides for use as vaccines. Such immunogenic polypeptides are described above and summarized in Table 4, below.

As used herein, a "related organism" is a broad term which refers to any organism whose growth or pathogenicity can be modulated by one of the pharmaceutical agents of the present invention. In general, such an organism will contain a homolog of the protein which is the target of the pharmaceutical agent or the protein used as a vaccine. As such, related organisms do not need to be bacterial but may be fungal or viral pathogens.

The pharmaceutical agents and compositions of the present invention may be administered in a convenient manner, such as by the oral, topical, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes. The pharmaceutical compositions are administered in an amount which is effective for treating and/or prophylaxis of the specific indication. In general, they are administered in an amount of at least about 1 mg/kg body weight and in most cases they will be administered in an amount not in excess of about 1 g/kg body weight per day. In most cases, the dosage is from about 0.1 mg/kg to about 10 g/kg body weight daily, taking into account the routes of administration, symptoms, etc.

The agents of the present invention can be used in native form or can be modified to form a chemical derivative. As used herein, a molecule is said to be a "chemical derivative" of another molecule when it contains additional chemical moieties not normally a part of the molecule. Such moieties may improve the molecule's solubility, absorption, biological half life, etc. The moieties may alternatively decrease the toxicity of the molecule, eliminate or attenuate any undesirable side effect of the molecule, etc. Moieties capable of mediating such effects are disclosed in, among other sources, REMINGTON'S PHARMACEUTICAL SCIENCES (1980) cited elsewhere herein.

For example, such moieties may change an immunological character of the functional derivative, such as affinity for a given antibody. Such changes in immunomodulation activity are measured by the appropriate assay, such as a competitive type immunoassay. Modifications of such protein properties as redox or thermal stability, biological half-life, hydrophobicity, susceptibility to proteolytic degradation or the tendency to aggregate with carriers or into multimers also may be effected in this way and can be assayed by methods well known to the skilled artisan.

The therapeutic effects of the agents of the present invention may be obtained by providing the agent to a patient by any suitable means (e.g., inhalation, intravenously, intramuscularly, subcutaneously, enterally, or parenterally). It is preferred to administer the agent of the present invention so as to achieve an effective concentration within the blood or tissue in which the growth of the organism is to be controlled. To achieve an effective blood concentration, the preferred method is to administer the agent by injection. The administration may be by continuous infusion, or by single or multiple injections.

In providing a patient with one of the agents of the present invention, the dosage of the administered agent will vary depending upon such factors as the patient's age, weight, height, sex, general medical condition, previous medical history, etc. In general, it is desirable to provide the recipient with a dosage of agent which is in the range of from about 1 pg/kg to 10 mg/kg (body weight of patient) although a lower or higher dosage may be administered. The therapeutically effective dose can be lowered by using combinations of the agents of the present invention or another agent.

As used herein, two or more compounds or agents are said to be administered "in combination" with each other when either (1) the physiological effects of each compound, or (2) the serum concentrations of each compound can

be measured at the same time. The composition of the present invention can be administered concurrently with, prior to, or following the administration of the other agent.

The agents of the present invention are intended to be provided to recipient subjects in an amount sufficient to decrease the rate of growth (as defined above) of the target organism.

The administration of the agent(s) of the invention may be for either a "prophylactic" or "therapeutic" purpose. When provided prophylactically, the agent(s) are provided in advance of any symptoms indicative of the organism's growth. The prophylactic administration of the agent(s) serves to prevent, attenuate, or decrease the rate of onset of any subsequent infection. When provided therapeutically, the agent(s) are provided at (or shortly after) the onset of an indication of infection. The therapeutic administration of the compound(s) serves to attenuate the pathological symptoms of the infection and to increase the rate of recovery.

The agents of the present invention are administered to a subject, such as a mammal, or a patient, in a pharmaceutically acceptable form and in a therapeutically effective concentration. A composition is said to be "pharmacologically acceptable" if its administration can be tolerated by a recipient patient. Such an agent is said to be administered in a "therapeutically effective amount" if the amount administered is physiologically significant. An agent is physiologically significant if its presence results in a detectable change in the physiology of a recipient patient.

The agents of the present invention can be formulated according to known methods to prepare pharmaceutically useful compositions, whereby these materials, or their functional derivatives, are combined in admixture with a pharmaceutically acceptable carrier vehicle. Suitable vehicles and their formulation, inclusive of other human proteins, e.g., human serum albumin, are described, for example, in REMINGTON'S PHARMACEUTICAL SCIENCES, 16<sup>th</sup> Ed., Osol, A., Ed., Mack Publishing, Easton PA (1980). In order to form a pharmaceutically acceptable composition suitable for effective administration, such compositions will contain an effective amount of one or more of the agents of the present invention, together with a suitable amount of carrier vehicle.

Additional pharmaceutical methods may be employed to control the duration of action. Control release preparations may be achieved through the use of polymers to complex or absorb one or more of the agents of the present invention. The controlled delivery may be effectuated by a variety of well known techniques, including formulation with macromolecules such as, for example, polyesters, polyamino acids, polyvinyl, pyrrolidone, ethylenevinylacetate, methylcellulose, carboxymethylcellulose, or protamine, sulfate, adjusting the concentration of the macromolecules and the agent in the formulation, and by appropriate use of methods of incorporation, which can be manipulated to effectuate a desired time course of release. Another possible method to control the duration of action by controlled release preparations is to incorporate agents of the present invention into particles of a polymeric material such as polyesters, polyamino acids, hydrogels, poly(lactic acid) or ethylene vinylacetate copolymers. Alternatively, instead of incorporating these agents into polymeric particles, it is possible to entrap these materials in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization with, for example, hydroxymethylcellulose or gelatine-microcapsules and poly(methylmethacrylate) microcapsules, respectively, or in colloidal drug delivery systems, for example liposomes, albumin microspheres, microemulsions, nanoparticles, and nanocapsules or in macroemulsions. Such techniques are disclosed in REMINGTON'S PHARMACEUTICAL SCIENCES (1980).

The invention further provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration.

In addition, the agents of the present invention may be employed in conjunction with other therapeutic compounds.

## 6. Shot-Gun Approach to Megabase DNA Sequencing

The present invention further demonstrates that a large sequence can be sequenced using a random shotgun approach. This procedure, described in detail in the examples that follow, has eliminated the up front cost of isolating and ordering overlapping or contiguous subclones prior to the start of the sequencing protocols.

Certain aspects of the present invention are described in greater detail in the examples that follow. The examples are provided by way of illustration. Other aspects and embodiments of the present invention are contemplated by the inventors, as will be clear to those of skill in the art from reading the present disclosure.

## ILLUSTRATIVE EXAMPLES

## LIBRARIES AND SEQUENCING

## 1. Shotgun Sequencing Probability Analysis

The overall strategy for a shotgun approach to whole genome sequencing follows from the Lander and Waterman (Landerman and Waterman, *Genomics* 2: 231 (1988)) application of the equation for the Poisson distribution. According to this treatment, the probability,  $P_0$ , that any given base in a sequence of size  $L$ , in nucleotides, is not sequenced after a certain amount,  $n$ , in nucleotides, of random sequence has been determined can be calculated by the equation  $P_0 = e^{-m}$ , where  $m$  is  $L/n$ , the fold coverage. For instance, for a genome of 2.8 Mb,  $m=1$  when 2.8 Mb of sequence has been randomly generated (1X coverage). At that point,  $P_0 = e^{-1} = 0.37$ . The probability that any given base has not been sequenced is the same as the probability that any region of the whole sequence  $L$  has not been determined and, therefore, is equivalent to the fraction of the whole sequence that has yet to be determined. Thus, at one-fold coverage, approximately 37% of a polynucleotide of size  $L$ , in nucleotides has not been sequenced. When 14 Mb of sequence has been generated, coverage is 5X for a 2.8 Mb and the unsequenced fraction drops to .0067 or 0.67%. 5X coverage of a 2.8 Mb sequence can be attained by sequencing approximately 17,000 random clones from both insert ends with an average sequence read length of 410 bp.

Similarly, the total gap length,  $G$ , is determined by the equation  $G = Le^{-m}$ , and the average gap size,  $g$ , follows the equation,  $g = L/n$ . Thus, 5X coverage leaves about 240 gaps averaging about 82 bp in size in a sequence of a polynucleotide 2.8 Mb long.

The treatment above is essentially that of Lander and Waterman, *Genomics* 2: 231 (1988)

## 2. Random Library Construction

In order to approximate the random model described above during actual sequencing, a nearly ideal library of cloned genomic fragments is required. The following library construction procedure was developed to achieve this end.

*Staphylococcus aureus* DNA was prepared by phenol extraction. A mixture containing 600  $\mu$ g DNA in 3.3 ml of 300 mM sodium acetate, 10 mM Tris-HCl, 1 mM Na-EDTA, 30% glycerol was sonicated for 1 min. at 0°C in a Branson Model 450 Sonicator at the lowest energy setting using a 3 mm probe. The sonicated DNA was ethanol precipitated and redissolved in 500  $\mu$ l TE buffer.

To create blunt-ends, a 100  $\mu$ l aliquot of the resuspended DNA was digested with 5 units of BAL31 nuclease (New England BioLabs) for 10 min at 30°C in 200  $\mu$ l BAL31 buffer. The digested DNA was phenol-extracted, ethanol-precipitated, redissolved in 100  $\mu$ l TE buffer, and then size-fractionated by electrophoresis through a 1.0% low melting temperature agarose gel. The section containing DNA fragments 1.6-2.0 kb in size was excised from the gel, and the LGT agarose was melted and the resulting solution was extracted with phenol to separate the agarose from the DNA. DNA was ethanol precipitated and redissolved in 20  $\mu$ l of TE buffer for ligation to vector.

A two-step ligation procedure was used to produce a plasmid library with 97% inserts, of which >99% were single inserts. The first ligation mixture (50  $\mu$ l) contained 2  $\mu$ g of DNA fragments, 2  $\mu$ g pUC18 DNA (Pharmacia) cut with SmaI and dephosphorylated with bacterial alkaline phosphatase, and 10 units of T4 ligase (GIBCO/BRL) and was incubated at 14°C for 4 hr. The ligation mixture then was phenol extracted and ethanol precipitated, and the precipitated DNA was dissolved in 20  $\mu$ l TE buffer and electrophoresed on a 1.0% low melting agarose gel. Discrete bands in a ladder were visualized by ethidium bromide-staining and UV illumination and identified by size as insert (i), vector (v), v+i, v+2i, v+3i, etc. The portion of the gel containing v+i DNA was excised and the v+i DNA was recovered and resuspended into 20  $\mu$ l TE. The v+i DNA then was blunt-ended by T4 polymerase treatment for 5 min. at 37°C in a reaction mixture (50  $\mu$ l) containing the v+i linears, 500  $\mu$ M each of the 4 dNTPs, and 9 units of T4 polymerase (New England BioLabs), under recommended buffer conditions. After phenol extraction and ethanol precipitation the repaired v+i linears were dissolved in 20  $\mu$ l TE. The final ligation to produce circles was carried out in a 50  $\mu$ l reaction containing 5  $\mu$ l of v+i linears and 5 units of T4 ligase at 14°C overnight. After 10 min. at 70°C the following day, the reaction mixture was stored at -20°C.

This two-stage procedure resulted in a molecularly random collection of single-insert plasmid recombinants with minimal contamination from double-insert chimeras (<1%) or free vector (<3%).

Since deviation from randomness can arise from propagation of the DNA in the host, *E. coli* host cells deficient in all recombination and restriction functions (A. Greener, *Strategies* 3 (1) 5 (1990)) were used to prevent rearrangements, deletions, and loss of clones by restriction. Furthermore, transformed cells were plated directly on antibiotic diffusion plates to avoid the usual broth recovery phase which allows multiplication and selection of the most rapidly growing cells.

Plating was carried out as follows. A 100  $\mu$ l aliquot of Epicurian Coli SURE II Supercompetent Cells (Stratagene 200152) was thawed on ice and transferred to a chilled Falcon 2059 tube on ice. A 1.7  $\mu$ l aliquot of 1.42 M beta-

mercaptoethanol was added to the aliquot of cells to a final concentration of 25 mM. Cells were incubated on ice for 10 min. A 1 µl aliquot of the final ligation was added to the cells and incubated on ice for 30 min. The cells were heat pulsed for 30 sec. at 42° C and placed back on ice for 2 min. The outgrowth period in liquid culture was eliminated from this protocol in order to minimize the preferential growth of any given transformed cell. Instead the transformation mixture was plated directly on a nutrient rich SOB plate containing a 5 ml bottom layer of SOB agar (5% SOB agar, 20 g tryptone, 5 g yeast extract, 0.5 g NaCl, 1.5% Difco Agar per liter of media). The 5 ml bottom layer is supplemented with 0.4 ml of 50 mg/ml ampicillin per 100 ml SOB agar. The 15 ml top layer of SOB agar is supplemented with 1 ml X-Gal (2%), 1 ml MgCl<sub>2</sub> (1 M) and 1 ml MgSO<sub>4</sub>/100 ml SOB agar. The 15 ml top layer was poured just prior to plating. Our titer was approximately 100 colonies/10 µl aliquot of transformation.

All colonies were picked for template preparation regardless of size. Thus, only clones lost due to "poison" DNA or deleterious gene products would be deleted from the library, resulting in a slight increase in gap number over that expected.

### 3. Random DNA Sequencing

High quality double stranded DNA plasmid templates were prepared using an alkaline lysis method developed in collaboration with 5Prime → 3Prime Inc. (Boulder, CO). Plasmid preparation was performed in a 96-well format for all stages of DNA preparation from bacterial growth through final DNA purification. Average template concentration was determined by running 25% of the samples on an agarose gel. DNA concentrations were not adjusted.

Templates were also prepared from a *Staphylococcus aureus* lambda genomic library. An unamplified library was constructed in Lambda DASH II vector (Stratagene). *Staphylococcus aureus* DNA (> 100 kb) was partially digested in a reaction mixture (200 µl) containing 50 µg DNA, 1X Sau3AI buffer, 20 units Sau3AI for 6 min. at 23 C. The digested DNA was phenol-extracted and centrifuges over a 10- 40% sucrose gradient. Fractions containing genomic DNA of 15-25 kb were recovered by precipitation. One µl of fragments was used with 1 µl of DASHII vector (Stratagene) in the recommended ligation reaction. One µl of the ligation mixture was used per packaging reaction following the recommended protocol with the Gigapack II XL Packaging Extract. Phage were plated directly without amplification from the packaging mixture (after dilution with 500 µl of recommended SM buffer and chloroform treatment). Yield was about 2.5x10<sup>9</sup> pfu/µl.

An amplified library was prepared from the primary packaging mixture according to the manufacturer's protocol. The amplified library is stored frozen in 7% dimethylsulfoxide. The phage titer is approximately 1x10<sup>9</sup> pfu/ml.

Mini-liquid lysates (0.1 µl) are prepared from randomly selected plaques and template is prepared by long range PCR. Samples are PCR amplified using modified T3 and T7 primers, and Elongase Supermix (LTI).

Sequencing reactions are carried out on plasmid templates using a combination of two workstations (BIOMEK 1000 and Hamilton Microlab 2200) and the Perkin-Elmer 9600 thermocycler with Applied Biosystems PRISM Ready Reaction Dye Primer Cycle Sequencing Kits for the M13 forward (M13-21) and the M13 reverse (M13RP1) primers. Dye terminator sequencing reactions are carried out on the lambda templates on a Perkin-Elmer 9600 Thermocycler using the Applied Biosystems Ready Reaction Dye Terminator Cycle Sequencing kits. Modified T7 and T3 primers are used to sequence the ends of the inserts from the Lambda DASH II library. Sequencing reactions are on a combination of ABI 373 DNA Sequencers and ABI 377 DNA sequencers. All of the dye terminator sequencing reactions are analyzed using the 2X 9 hour module on the AB 377. Dye primer reactions are analyzed on a combination of ABI 373 and ABI 377 DNA sequencers. The overall sequencing success rate very approximately is about 85% for M13-21 and M13RP1 sequences and 65% for dye-terminator reactions. The average usable read length is 485 bp for M13-21 sequences, 445bp for M13RP1 sequences, and 375 bp for dye-terminator reactions.

### 4. Protocol for Automated Cycle Sequencing

The sequencing was carried out using Hamilton Microstation 2200, Perkin Elmer 9600 thermocyclers, ABI 373 and ABI 377 Automated DNA Sequencers. The Hamilton combines pre-aliquoted templates and reaction mixes consisting of deoxy- and dideoxynucleotides, the thermostable Taq DNA polymerase, fluorescently-labelled sequencing primers, and reaction buffer. Reaction mixes and templates were combined in the wells of a 96-well thermocycling plate and transferred to the Perkin Elmer 9600 thermocycler. Thirty consecutive cycles of linear amplification (i.e. one primer synthesis) steps were performed including denaturation, annealing of primer and template, and extension, i.e. DNA synthesis. A heated lid with rubber gaskets on the thermocycling plate prevents evaporation without the need for an oil overlay.

Two sequencing protocols were used: one for dye-labelled primers and a second for dye-labelled dideoxy chain terminators. The shotgun sequencing involves use of four dye-labelled sequencing primers, one for each of the four terminator nucleotide. Each dye-primer was labelled with a different fluorescent dye, permitting the four individual reactions to be combined into one lane of the 373 or 377 DNA Sequencer for electrophoresis, detection, and base-

calling. ABI currently supplies premixed reaction mixes in bulk packages containing all the necessary non-template reagents for sequencing. Sequencing can be done with both plasmid and PCR-generated templates with both dye-primers and dye-terminators with approximately equal fidelity, although plasmid templates generally give longer usable sequences.

Thirty-two reactions were loaded per ABI 373 Sequencer each day and 96 samples can be loaded on an ABI 377 per day. Electrophoresis was run overnight (ABI 373) or for 2 1/2 hours (ABI 377) following the manufacturer's protocols. Following electrophoresis and fluorescence detection, the ABI 373 or ABI 377 performs automatic lane tracking and base-calling. The lane-tracking was confirmed visually. Each sequence electropherogram (or fluorescence lane trace) was inspected visually and assessed for quality. Trailing sequences of low quality were removed and the sequence itself was loaded via software to a Sybase database (archived daily to 8mm tape). Leading vector polylinker sequence was removed automatically by a software program. Average edited lengths of sequences from the standard ABI 373 or ABI 377 were around 400 bp and depend mostly on the quality of the template used for the sequencing reaction.

## INFORMATICS

### 1. Data Management

A number of information management systems for a large-scale sequencing lab have been developed. (For review see, for instance, Kerlavage *et al.*, *Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System Sciences*, IEEE Computer Society Press, Washington D. C., 585 (1993)) The system used to collect and assemble the sequence data was developed using the Sybase relational database management system and was designed to automate data flow wherever possible and to reduce user error. The database stores and correlates all information collected during the entire operation from template preparation to final analysis of the genome. Because the raw output of the ABI 373 Sequencers was based on a Macintosh platform and the data management system chosen was based on a Unix platform, it was necessary to design and implement a variety of multi-user, client-server applications which allow the raw data as well as analysis results to flow seamlessly into the database with a minimum of user effort.

### 2. Assembly

An assembly engine (TIGR Assembler) developed for the rapid and accurate assembly of thousands of sequence fragments was employed to generate contigs. The TIGR assembler simultaneously clusters and assembles fragments of the genome. In order to obtain the speed necessary to assemble more than  $10^4$  fragments, the algorithm builds a hash table of 12 bp oligonucleotide subsequences to generate a list of potential sequence fragment overlaps. The number of potential overlaps for each fragment determines which fragments are likely to fall into repetitive elements. Beginning with a single seed sequence fragment, TIGR Assembler extends the current contig by attempting to add the best matching fragment based on oligonucleotide content. The contig and candidate fragment are aligned using a modified version of the Smith-Waterman algorithm which provides for optimal gapped alignments (Waterman, M. S., *Methods in Enzymology* 164: 765 (1988)). The contig is extended by the fragment only if strict criteria for the quality of the match are met. The match criteria include the minimum length of overlap, the maximum length of an unmatched end, and the minimum percentage match. These criteria are automatically lowered by the algorithm in regions of minimal coverage and raised in regions with a possible repetitive element. The number of potential overlaps for each fragment determines which fragments are likely to fall into repetitive elements. Fragments representing the boundaries of repetitive elements and potentially chimeric fragments are often rejected based on partial mismatches at the ends of alignments and excluded from the current contig. TIGR Assembler is designed to take advantage of clone size information coupled with sequencing from both ends of each template. It enforces the constraint that sequence fragments from two ends of the same template point toward one another in the contig and are located within a certain ranged of base pairs (definable for each clone based on the known clone size range for a given library).

### 3. Identifying Genes

The predicted coding regions of the *Staphylococcus aureus* genome were initially defined with the program zorf, which finds ORFs of a minimum length. The predicted coding region sequences were used in searches against a database of all *Staphylococcus aureus* nucleotide sequences from GenBank (release 92.0), using the BLASTN search method to identify overlaps of 50 or more nucleotides with at least a 95% identity. Those ORFs with nucleotide sequence matches are shown in Table 1. The ORFs without such matches were translated to protein sequences and compared to a non-redundant database of known proteins generated by combining the Swiss-prot, PIR and GenPept databases. ORFs of at least 80 amino acids that matched a database protein with BLASTP probability less than or equal to 0.01 are shown in Table 2. The table also lists assigned functions based on the closest match in the databases.

ORFs of at least 120 amino acids that did not match protein or nucleotide sequences in the databases at these levels are shown in Table 3

## ILLUSTRATIVE APPLICATIONS

### 1. Production of an Antibody to a *Staphylococcus aureus* Protein

Substantially pure protein or polypeptide is isolated from the transfected or transformed cells using any one of the methods known in the art. The protein can also be produced in a recombinant prokaryotic expression system, such as *E. coli*, or can be chemically synthesized. Concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms/ml. Monoclonal or polyclonal antibody to the protein can then be prepared as follows:

### 2. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes of any of the peptides identified and isolated as described can be prepared from murine hybridomas according to the classical method of Kohler, G., and Milstein, C., Nature 256:495 (1975) or modifications of the methods thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the selected protein over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, E., Meth. Enzymol. 70:419 (1980), and modified methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. *et al.* Basic Methods in Molecular Biology Elsevier, New York, Section 21-2 (1989).

### 3. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogenous epitopes of a single protein can be prepared by immunizing suitable animals with the expressed protein described above, which can be unmodified or modified to enhance immunogenicity. Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. For example, small molecules tend to be less immunogenic than other and may require the use of carriers and adjuvant. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. An effective immunization protocol for rabbits can be found in Vaitukaitis, J. *et al.*, J. Clin. Endocrinol. Metab. 33:988-991 (1971).

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, O. *et al.*, Chap. 19 in Handbook of Experimental Immunology, Wier, D., ed, Blackwell (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12M). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, D., Chap. 42 in Manual of Clinical Immunology, second edition, Rose and Friedman, eds., Amer. Soc. For Microbiology, Washington, D. C. (1980).

Antibody preparations prepared according to either protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample. In addition, they are useful in various animal models of Staphylococcal disease known to those of skill in the art as a means of evaluating the protein used to make the antibody as a potential vaccine target or as a means of evaluating the antibody as a potential immunotherapeutic reagent.

### 3. Preparation of PCR Primers and Amplification of DNA

Various fragments of the *Staphylococcus aureus* genome, such as those of Tables 1-3 and SEQ ID NOS: 1-5, 191 can be used, in accordance with the present invention, to prepare PCR primers for a variety of uses. The PCR primers are preferably at least 15 bases, and more preferably at least 18 bases in length. When selecting a primer sequence, it is preferred that the primer pairs have approximately the same G/C ratio, so that melting temperatures are approxi-

mately the same. The PCR primers and amplified DNA of this Example find use in the Examples that follow.

#### 4. Gene expression from DNA Sequences Corresponding to ORFs

A fragment of the *Staphylococcus aureus* genome provided in Tables 1-3 is introduced into an expression vector using conventional technology. Techniques to transfer cloned sequences into expression vectors that direct protein translation in mammalian, yeast, insect or bacterial expression systems are well known in the art. Commercially available vectors and expression systems are available from a variety of suppliers including Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence may be optimized for the particular expression organism, as explained by Hatfield *et al.*, U. S. Patent No. 5,082,767, incorporated herein by this reference.

The following is provided as one exemplary method to generate polypeptide(s) from cloned ORFs of the *Staphylococcus aureus* genome fragment. Bacterial ORFs generally lack a poly A addition signal. The addition signal sequence can be added to the construct by, for example, splicing out the poly A addition sequence from pSG5 (Stratagene) using BglII and SalI restriction endonuclease enzymes and incorporating it into the mammalian expression vector pXT1 (Stratagene) for use in eukaryotic expression systems. pXT1 contains the LTRs and a portion of the gag gene of Moloney Murine Leukemia Virus. The positions of the LTRs in the construct allow efficient stable transfection. The vector includes the Herpes Simplex thymidine kinase promoter and the selectable neomycin gene. The *Staphylococcus aureus* DNA is obtained by PCR from the bacterial vector using oligonucleotide primers complementary to the *Staphylococcus aureus* DNA and containing restriction endonuclease sequences for PstI incorporated into the 5' primer and BglII at the 5' end of the corresponding *Staphylococcus aureus* DNA 3' primer, taking care to ensure that the *Staphylococcus aureus* DNA is positioned such that its followed with the poly A addition sequence. The purified fragment obtained from the resulting PCR reaction is digested with PstI, blunt ended with an exonuclease, digested with BglII, purified and ligated to pXT1, now containing a poly A addition sequence and digested BglII.

The ligated product is transfected into mouse NIH 3T3 cells using Lipofectin (Life Technologies, Inc., Grand Island, New York) under conditions outlined in the product specification. Positive transfectants are selected after growing the transfected cells in 600 ug/ml G418 (Sigma, St. Louis, Missouri). The protein is preferably released into the supernatant. However if the protein has membrane binding domains, the protein may additionally be retained within the cell or expression may be restricted to the cell surface. Since it may be necessary to purify and locate the transfected product, synthetic 15-mer peptides synthesized from the predicted *Staphylococcus aureus* DNA sequence are injected into mice to generate antibody to the polypeptide encoded by the *Staphylococcus aureus* DNA.

Alternatively and if antibody production is not possible, the *Staphylococcus aureus* DNA sequence is additionally incorporated into eukaryotic expression vectors and expressed as, for example, a globin fusion. Antibody to the globin moiety then is used to purify the chimeric protein. Corresponding protease cleavage sites are engineered between the globin moiety and the polypeptide encoded by the *Staphylococcus aureus* DNA so that the latter may be freed from the formed by simple protease digestion. One useful expression vector for generating globin chimerics is pSG5 (Stratagene). This vector encodes a rabbit globin. Intron II of the rabbit globin gene facilitates splicing of the expressed transcript, and the polyadenylation signal incorporated into the construct increases the level of expression. These techniques are well known to those skilled in the art of molecular biology. Standard methods are published in methods texts such as Davis *et al.*, cited elsewhere herein, and many of the methods are available from the technical assistance representatives from Stratagene, Life Technologies, Inc., or Promega. Polypeptides of the invention also may be produced using *in vitro* translation systems such as *in vitro* Express™ Translation Kit (Stratagene).

While the present invention has been described in some detail for purposes of clarity and understanding, one skilled in the art will appreciate that various changes in form and detail can be made without departing from the true scope of the invention.

All patents, patent applications and publications referred to above are hereby incorporated by reference.



S. aureus - Coding regions containing known sequences

Contig ID	ICRF ID	Start (nt)	Stop (nt)	match accession	match gene name	Percent ident	HSP nt length	ORF nt length
1	1	1419	757	[emb]X17301[CAHD]	S. aureus DNA for hid gene and for part of agr gene	100	663	663
1	2	3273	2452	[emb]X52543[SAAG]	S. aureus agrA, agrB and hid genes	99	809	822
1	5	6418	5651	[dbj]D147111[STAH]	Staphylococcus aureus HSP10 and HSP60 genes	98	223	768
5	1	807	439	[emb]X72700[SAPV]	S. aureus genes for S and F components of Pantone-Valentine leucocidin	81	216	369
5	4	5031	3571	[emb]X72700[SAPV]	S. aureus genes for S and F components of Pantone-Valentine leucocidin	95	424	1461
10	1	86	904	[gb]L25288[ ]	Staphylococcus aureus gyrase-like protein alpha and beta subunit (grIA and grIB) genes, complete cds	98	715	819
16	5	5302	6246	[gb]U35773[ ]	Staphylococcus aureus prolipoprotein diacylglycerol transferase (lgt) gene, complete cds	94	251	945
16	6	6249	7091	[gb]U35773[ ]	Staphylococcus aureus prolipoprotein diacylglycerol transferase (lgt) gene, complete cds	99	843	843
16	7	7084	7584	[gb]U35773[ ]	Staphylococcus aureus prolipoprotein diacylglycerol transferase (lgt) gene, complete cds	99	342	501
20	1	995	549	[gb]U19300[ ]	Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank	100	443	447
20	2	1011	841	[gb]U19300[ ]	Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank	91	137	171
20	3	2010	1798	[gb]U19300[ ]	Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank	100	110	213
20	4	5300	3825	[gb]M16714[ ]	Staphylococcus aureus peptidoglycan hydrolase gene, complete cds	100	948	1476
20	5	4788	4282	[gb]M16714[ ]	Staphylococcus aureus peptidoglycan hydrolase gene, complete cds	100	309	507
26	1	2	145	[gb]U41072[ ]	Staphylococcus aureus isoleucyl-tRNA synthetase (ileS) gene, partial cds	100	126	144
26	2	84	557	[gb]U41072[ ]	Staphylococcus aureus isoleucyl-tRNA synthetase (ileS) gene, partial cds	99	430	474
26	3	763	3531	[emb]X74219[SAIL]	S. aureus gene for isoleucyl-tRNA synthetase	99	2769	2769
29	3	1261	4392	[gb]U6665[ ]	Staphylococcus aureus DNA fragment with class II promoter activity	100	117	332
31	14	14977	13463	[emb]X73889[SAP1]	S. aureus genes P1 and P2	99	1351	1515
31	15	14241	13855	[emb]X73889[SAP1]	S. aureus genes P1 and P2	98	258	387
38	17	14284	13112	[gb]M12715[ ]	S. aureus gdh gene encoding lipase (glycerol ester hydrolase)	100	372	1173
38	19	13434	13518	[gb]M12715[ ]	S. aureus gdh gene encoding lipase (glycerol ester hydrolase)	100	2085	2085

TABLE I

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	Percent ident	HSP nt length	ORF nt length
46	3	1720	1727	gb U73374	Staphylococcus aureus type 8 capsule genes, cap8A, cap8U, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	98	1209	1709
46	3	1720	2295	gb U73374	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	98	576	576
46	4	2259	3182	gb U73374	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	97	924	924
46	5	3173	4498	gb U73374	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	98	1283	1326
46	6	4536	5720	gb U73374	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	98	1185	1185
46	7	6455	6120	gb U73374	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	99	278	336
48	1	2	955	gb U25893	Staphylococcus aureus recA gene, complete cds	99	954	954
50	3	4465	2924	emb X65029 SAAH	S. aureus AhpC gene	100	88	1542
50	4	4108	3515	emb X65029 SAAH	S. aureus AhpC gene	98	540	594
54	3	5074	3392	emb X62592 SAFN	S. aureus fnbB gene for fibronectin binding protein B	100	1668	1683
54	4	4865	4122	emb X62592 SAFN	S. aureus fnbB gene for fibronectin binding protein B	99	720	744
54	5	5056	4562	emb X62592 SAFN	S. aureus fnbB gene for fibronectin binding protein B	100	463	495
54	6	11386	8300	gb J04151	S. aureus fibronectin-binding protein (fnbA) mRNA, complete cds	100	3087	3087
58	3	1743	2819	emb X87104 SADN	S. aureus mdr, pbp4 and tagD genes (SGS11-55 isolate)	89	68	1077
58	4	2858	3280	emb X91786 SAPB	S. aureus abcA, pbp4, and tagD genes	99	423	423
58	5	6005	4701	emb X91786 SAPB	S. aureus abcA, pbp4, and tagD genes	99	1305	1305
58	6	5677	5378	gb U29478	Staphylococcus aureus ABC transporter-like protein AbcA (abcA) gene, partial cds	100	300	300
58	7	5086	6840	emb X91786 SAPB	S. aureus abcA, pbp4, and tagD genes	99	1755	1755
72	1	888	445	gb U21854	S. aureus agr gene encoding an accessory gene regulator protein, complete cds	100	444	444
72	2	2457	1453	emb X52543 SAAG	S. aureus agrA, agrB and hld genes	99	873	1005

TABLE 1

S aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	Percent ident	HSP length	ORF length
82	1	357	3917	emb[X64172]SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF203, DNA-directed RNA polymerase beta & beta' chains	99	2396	3561
82	2	4027	7677	emb[X89233]SARP	S.aureus DNA for rpoC gene	99	3171	3651
82	3	7745	8068	gb[U20869]	Staphylococcus aureus ribosomal protein S12 (rpsL) gene, complete cds, ribosomal protein S7 (rpsG) and ORF 1 genes, partial cds	100	320	324
82	4	8103	8579	gb[U20869]	Staphylococcus aureus ribosomal protein S12 (rpsL) gene, complete cds, ribosomal protein S7 (rpsG) and ORF 1 genes, partial cds	100	477	477
82	5	8618	8821	gb[U20869]	Staphylococcus aureus ribosomal protein S12 (rpsL) gene, complete cds, ribosomal protein S7 (rpsG) and ORF 1 genes, partial cds	100	154	204
84	1	18	191	gb[U73374]	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	98	164	174
84	2	189	893	gb[U73374]	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	94	705	705
84	3	887	1660	gb[U73374]	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	99	774	774
84	4	1584	3503	gb[U73374]	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	98	1920	1920
84	5	3394	4521	gb[U73374]	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	97	1128	1128
84	6	4519	5643	gb[U73374]	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	97	1125	1125
96	2	1245	3896	emb[Z18852]SACF	S.aureus gene for clumping factor	83	660	2652
97	2	625	882	gb[U41072]	Staphylococcus aureus isoleucyl-tRNA synthetase (ileS) gene, partial cds	97	68	258
111	1	3	452	gb[U41499]	Staphylococcus aureus ORF1, partial cds, ORF2, ORF3, autolysin (atl) genes, complete cds	100	450	450
111	2	526	1041	gb[U41499]	Staphylococcus aureus ORF1, partial cds, ORF2, ORF3, autolysin (atl) genes, complete cds	99	516	516
117	2	1278	1958	gb[M83994]	Staphylococcus aureus prolipoprotein signal peptidase (lsp) gene, complete cds	100	61	681

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	percent ident	HSP nt length	ORF nt length
118	4	3787	4254	[dbj]U30690 STAN	Staphylococcus aureus genes for ORF37, HSP20, HSP70, HSP10, ORF35, complete cds	99	467	468
130	4	2597	3640	[emb]X13290 SATN	Staphylococcus aureus multi-resistance plasmid pSK1 DNA containing transposon Tn4003	78	956	1044
130	5	3813	4265	[emb]Z16422 SADI	S. aureus dfrB gene for dihydrofolate reductase	98	416	453
130	6	4309	5172	[emb]Z16422 SADI	S. aureus dfrB gene for dihydrofolate reductase	98	607	864
136	4	5296	6207	[emb]X71437 SAGY	S. aureus genes gyrB, gyrA and recF (partial)	97	838	912
136	5	11680	8987	[dbj]U10489 STAG	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	100	2694	2694
136	6	12686	10940	[dbj]U10489 STAG	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	99	1947	1947
136	7	12592	11765	[gb]S77055	recF cluster: dnaA-replisome assembly protein... gyrB-DNA gyrase beta subunit [Staphylococcus aureus, YB886, Genomic, 5 genes, 3573 nt]	99	822	828
143	3	4371	2867	[gb]U36379	Staphylococcus aureus S-adenosylmethionine synthetase gene, complete cds	99	1305	1305
143	4	3100	4281	[gb]U42943	Staphylococcus aureus (clone KIN50) phosphoenolpyruvate carboxykinase (pckA) gene, complete cds	100	1170	1182
143	5	4254	4718	[gb]U51133	Staphylococcus aureus phosphoenolpyruvate carboxykinase (pckA) gene, complete cds	100	449	465
143	9	6977	7261	[gb]U51132	Staphylococcus aureus o-succinylbenzoic acid CoA ligase (meae), and o-succinylbenzoic acid synthetase (menc) genes, complete cds	100	75	285
143	10	9464	8361	[gb]U51132	Staphylococcus aureus o-succinylbenzoic acid CoA ligase (meae), and o-succinylbenzoic acid synthetase (menc) genes, complete cds	100	1104	1104
143	11	11232	9748	[gb]U51132	Staphylococcus aureus o-succinylbenzoic acid CoA ligase (meae), and o-succinylbenzoic acid synthetase (menc) genes, complete cds	100	1485	1485
143	12	10739	10320	[gb]U51132	Staphylococcus aureus o-succinylbenzoic acid CoA ligase (meae), and o-succinylbenzoic acid synthetase (menc) genes, complete cds	100	332	420
152	5	2454	3437	[emb]X58434 SAPD	S. aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydroliipoamide acetyltransferase and dihydroliipoamide dehydrogenase	99	305	984
152	6	3513	4820	[emb]X58434 SAPD	S. aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydroliipoamide acetyltransferase and dihydroliipoamide dehydrogenase	98	1308	1308
152	7	4818	6230	[emb]X58434 SAPD	S. aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydroliipoamide acetyltransferase and dihydroliipoamide dehydrogenase	99	1413	1413
153	1	387	1526	[gb]S77055	recF cluster: dnaA-replisome assembly protein... gyrB-DNA gyrase beta subunit [Staphylococcus aureus, YB886, Genomic, 5 genes, 3573 nt]	99	1140	1140
153	2	1877	2152	[gb]S77055	recF cluster: dnaA-replisome assembly protein... gyrB-DNA gyrase beta subunit [Staphylococcus aureus, YB886, Genomic, 5 genes, 3573 nt]	100	276	276

TABLE 1

S. aureus - Coding regions containing known sequences

Contig in	ORF to	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
153	3	2143	2289	[gb S77055]	tecF cluster, dnaA-replicative assembly protein...gyrB-DNA gyrase beta subunit [Staphylococcus aureus, YB886, Genomic, 5 genes, 3573 nt]	99	133	147
154	110	10792	9314	[gb U06451]	[Staphylococcus aureus proline permease homolog (putP) gene, complete cds]	91	154	1479
154	111	9935	9615	[gb U06451]	[Staphylococcus aureus proline permease homolog (putP) gene, complete cds]	99	229	321
154	112	9943	10167	[gb U06451]	[Staphylococcus aureus proline permease homolog (putP) gene, complete cds]	94	123	225
154	113	10089	11501	[gb U06451]	[Staphylococcus aureus proline permease homolog (putP) gene, complete cds]	99	1326	1413
159	2	2195	1212	[dbj U28879 STAP	[Staphylococcus aureus gene for penicillin-binding protein 1, complete cds]	100	71	984
161	3	2596	2270	[gb H83994]	[Staphylococcus aureus prolipoprotein signal peptidase (lsp) gene, complete cds]	92	203	327
162	1	1406	705	[gb U21221]	[Staphylococcus aureus hyaluronate lyase (hysA) gene, complete cds]	100	702	702
163	4	1263	1772	[gb U19770]	[Staphylococcus aureus pyrrolidone carboxyl peptidase (pcp) gene, complete cds]	96	127	510
164	7	4774	9117	[dbj D86727 D867	[Staphylococcus aureus DNA for DNA polymerase III, complete cds]	99	3470	4344
168	7	7448	6447	[gb U21636]	[Staphylococcus aureus cnp-binding-factor 1 (cbf1) and ORF X genes, complete cds]	100	1002	1002
168	8	9538	7961	[gb U21636]	[Staphylococcus aureus cnp-binding-factor 1 (cbf1) and ORF X genes, complete cds]	99	1159	1578
173	6	9240	7801	[gb J03479]	[S. aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds]	100	1443	1440
173	7	11252	9522	[gb J03479]	[S. aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds]	99	1731	1731
173	8	8285	8704	[gb J03479]	[S. aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds]	100	420	420
173	9	10168	9839	[gb J03479]	[S. aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds]	100	330	330
173	10	11815	10829	[emb X14827 SALA	[Staphylococcus aureus lacC and lacD genes]	100	987	987
173	11	11271	11174	[emb X14827 SALA	[Staphylococcus aureus lacC and lacD genes]	100	948	948
173	12	112838	12305	[gb H64724]	[S. aureus tagatase 6-phosphate isomerase gene, complete cds]	100	534	534
173	13	13243	12773	[gb H32103]	[Staphylococcus aureus lac repressor (lacR) gene, complete cds and lacA repressor (lacA) partial cds]	100	471	471
173	14	114633	13866	[gb H32103]	[Staphylococcus aureus lac repressor (lacR) gene, complete cds and lacA repressor (lacA) partial cds]	100	768	768

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	percent ident	ASP nt length	ORF nt length
178	1	2	655	[gb U52961]	Staphylococcus aureus holin-like protein LrgA (lrgA) and LrgB (lrgB) genes, complete cds	100	115	654
178	2	2201	1482	[gb U52961]	Staphylococcus aureus holin-like protein LrgA (lrgA) and LrgB (lrgB) genes, complete cds	100	720	720
178	3	2361	1909	[gb U52961]	Staphylococcus aureus holin-like protein LrgA (lrgA) and LrgB (lrgB) genes, complete cds	100	453	453
178	4	1551	1853	[gb U52961]	Staphylococcus aureus holin-like protein LrgA (lrgA) and LrgB (lrgB) genes, complete cds	100	303	303
178	5	3541	2777	[gb U42945]	Staphylococcus aureus lytS and lytR genes, complete cds	99	765	765
178	6	3294	3025	[gb U42945]	Staphylococcus aureus lytS and lytR genes, complete cds	99	270	270
181	1	1114	590	[gb H61177]	S. aureus sigma factor (plac) gene, complete cds	99	499	525
182	1	3	341	[emb X61307]	Staphylococcus aureus spa gene for protein A	98	277	339
182	2	690	2312	[gb J01786]	S. aureus spa gene coding for protein A, complete cds	97	1332	1623
182	3	5861	4251	[emb X61307]	Staphylococcus aureus spa gene for protein A	99	119	1611
185	1	3	824	[gb U11979]	Staphylococcus aureus chorismate synthase (arcC) and nucleoside diphosphate kinase (ndk) genes, complete cds, dehydrogenase synthase (arob) and geranylgeranyl pyrophosphate synthetase homolog (gerCC) genes, partial cds	90	132	822
191	3	841	2760	[emb X17679]	SACO Staphylococcus aureus coa gene for coagulase	99	1920	1920
191	4	2967	3143	[emb X16457]	SAST Staphylococcus aureus gene for staphylocoagulase	99	177	177
191	5	5768	4566	[emb X16457]	SAST Staphylococcus aureus gene for staphylocoagulase	99	250	1203
196	1	1741	872	[gb L36472]	Staphylococcus aureus lyxJ-1 RNA synthetase gene, complete cds, transfer RNA (tRNA) genes, 5S ribosomal RNA (5S rRNA) gene, 16S ribosomal RNA (16S rRNA) gene, 23S ribosomal RNA (23S rRNA) gene	99	870	870
198	3	1688	2011	[emb X93205]	SAPT S. aureus ptbH and ptbI genes	99	324	324
198	4	2005	2310	[emb X93205]	SAPT S. aureus ptbH and ptbI genes	97	304	306
202	1	163	1305	[emb X97985]	SA12 S. aureus orfs 1, 2, 3 & 4	99	1143	1143
202	2	1303	2175	[emb X73889]	SAP1 S. aureus genes P1 and P2	94	444	873
210	1	1114	1558	[dbj D17366]	STAA Staphylococcus aureus atl gene for autolysin, complete cds and other ORFs	99	1552	1557
210	2	2939	2232	[gb L41499]	Staphylococcus aureus ORF1, partial cds, ORF2, ORF3, autolysin (atl) genes, complete cds	99	684	708
214	11	7429	7770	[dbj D86240]	D862 Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds	96	157	342

TABLE I

S. aureus - coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
216	3	398	1318	emb X72700 SAPV	S. aureus genes for S and F components of Pantone-Valentine leucocidins	88	265	921
219	2	1810	1073	dbj D30690 STAN	Staphylococcus aureus genes for ORF37, HSP70, HSP40, ORF35, complete cds	100	80	738
219	3	2979	2035	dbj D30690 STAN	Staphylococcus aureus genes for ORF37, HSP70, HSP40, ORF35, complete cds	99	945	945
219	4	4359	3196	dbj D30690 STAN	Staphylococcus aureus genes for ORF37, HSP70, HSP40, ORF35, complete cds	99	1164	1164
219	5	7044	5176	dbj D30690 STAN	Staphylococcus aureus genes for ORF37, HSP70, HSP40, ORF35, complete cds	98	1860	1860
219	6	6557	5883	dbj D30690 STAN	Staphylococcus aureus genes for ORF37, HSP70, HSP40, ORF35, complete cds	99	675	675
219	7	6801	6134	dbj D30690 STAN	Staphylococcus aureus genes for ORF37, HSP70, HSP40, ORF35, complete cds	98	468	468
221	8	10816	10034	gb U19298	Staphylococcus aureus phosphatidylinositol-specific phospholipase C (pic) gene, complete cds	91	67	783
223	1	2855	1506	gb U07374	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	99	102	1350
234	1	2	1357	emb X97985 SA12	S. aureus orfs 1,2,3 & 4	100	176	1356
234	2	1694	2485	emb X97985 SA12	S. aureus orfs 1,2,3 & 4	100	782	792
234	3	2648	3146	emb X97985 SA12	S. aureus orfs 1,2,3 & 4	99	501	501
234	4	3120	4604	emb X97985 SA12	S. aureus orfs 1,2,3 & 4	99	1303	1485
236	6	3826	5322	gb U48826	Staphylococcus aureus elastin binding protein (ebpS) gene, complete cds	96	648	1497
248	1	2	403	emb X62288 SAPE	S. aureus DNA for penicillin-binding protein 2	100	103	402
248	2	348	852	gb U35426	Staphylococcus aureus penicillin-binding protein 2 (pbp2) gene, complete cds	99	465	465
253	2	1539	1093	gb U46541	Staphylococcus aureus sarA gene, complete cds	96	447	447
254	2	150	1835	gb U57063	Staphylococcus aureus scdA gene, complete cds	94	142	1686
254	3	1973	2728	gb U57063	Staphylococcus aureus scdA gene, complete cds	99	756	756
260	1	2	1900	gb U90643	Staphylococcus aureus glycerol ester hydrolase (lil) gene, complete cds	90	1213	1499
265	1	1	942	db D21131 STAS	Staphylococcus aureus gene for a participant in homogeneous expression of high-level methicillin resistance, complete cds	99	941	942

TABLE I

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	Percent ident	ORF nt length
265	2	648	476	dbj 021131 STAS	Staphylococcus aureus gene for a participant in homogeneous expression of high-level methicillin resistance, complete cds	99	213
265	3	2418	1765	dbj 021131 STAS	Staphylococcus aureus gene for a participant in homogeneous expression of high-level methicillin resistance, complete cds	98	69
266	1	2	1014	dbj 014711 STAH	Staphylococcus aureus HSP10 and HSP60 genes	98	743
282	1	1	525	gb F72488	hemB-porphobilinogen synthase [Staphylococcus aureus, SA1959, Genomic, 1087 nt]	100	110
282	2	516	1502	gb F72488	hemB-porphobilinogen synthase [Staphylococcus aureus, SA1959, Genomic, 1087 nt]	100	952
284	1	3	170	gb H63176	Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds	98	84
284	2	282	1034	gb H63176	Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds	100	712
284	3	1028	2026	gb H63176	Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds	99	979
284	4	1990	2202	gb H63176	Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds	98	187
289	3	1536	1991	gb H32470	S aureus SauJAI-restriction-enzyme and SauJAI-modification-enzyme genes, complete cds	99	338
303	1	2	868	gb L01055	Staphylococcus aureus gamma-hemolysin components A, B and C (hlgA, hlgB, hlgC) genes, complete cds	99	867
303	2	1409	2383	gb L01055	Staphylococcus aureus gamma-hemolysin components A, B and C (hlgA, hlgB, hlgC) genes, complete cds	100	975
303	3	2367	3161	gb L01055	Staphylococcus aureus gamma-hemolysin components A, B and C (hlgA, hlgB, hlgC) genes, complete cds	99	793
305	1	2707	1355	dbj D17J66 STAA	Staphylococcus aureus atl gene for autolysin, complete cds and other ORFs	99	1343
311	1	2628	1315	gb L42945	Staphylococcus aureus lylS and lylR genes, complete cds	98	1314
312	6	7019	7870	gb L44017	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	74	351
323	1	1998	1003	gb U031175	Staphylococcus aureus D-specific D-2-hydroxyacid dehydrogenase (ddh) gene, complete cds	98	996
326	1	1	217	emb Y00156 SASP	Staphylococcus aureus V8 serine protease gene	100	108
338	1	687	388	emb X64389 SALE	S aureus leuF-p83 gene for F component of leucocidin R	98	259
338	2	1824	1048	emb Y04389 SALE	S aureus leuF-p83 gene for F component of leucocidin R	97	137

TABLE 1



S aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	Percent ident	ORF nt length
342	2	579	1754	[gb U06462]	Staphylococcus aureus SA4 Fls2 (fts2) gene, complete cds	100	1176
344	2	517	1248	[emb V01281 SANU]	S aureus mAna for nuclease	98	732
349	1	457	230	[gb M2039]	S aureus bacteriophage phi-11 attachment site (attB)	96	172
353	1	1016	516	[gb HM3994]	Staphylococcus aureus prolipoprotein signal peptidase (lsp) gene, complete cds	100	187
353	2	1582	1046	[gb HM3994]	Staphylococcus aureus prolipoprotein signal peptidase (lsp) gene, complete cds	99	537
356	1	3	674	[gb U20503]	Staphylococcus aureus MIC class II analog gene, complete cds	75	671
361	1	1	903	[gb U19298]	Staphylococcus aureus phosphatidylinositol-specific phospholipase C (plc) gene, complete cds	98	747
361	2	1103	1507	[gb U19298]	Staphylococcus aureus phosphatidylinositol-specific phospholipase C (plc) gene, complete cds	97	68
373	1	3	1148	[emb X62288 SAFE]	S aureus DNA for penicillin-binding protein 2	99	1146
389	3	1904	1248	[emb X62282 SAT5]	S aureus target site DNA for IS431 insertion	97	349
400	1	1	540	[emb X61716 SAIL]	S aureus hlb gene encoding sphingomyelinase	99	389
400	2	1693	1187	[emb X13404 SAIL]	Staphylococcus aureus hlb gene for beta-hemolysin	99	178
408	1	1810	1049	[gb S76213]	esp23-alkaline shock protein 23 (methicillin resistant) (Staphylococcus aureus 912 Genomic 1160 nt)	99	163
418	1	2	217	[gb U41499]	Staphylococcus aureus ORF1, partial cds, ORF2, ORF3, autolysin (atl) genes, complete cds	100	216
418	2	854	639	[dbj C17366 STAA]	Staphylococcus aureus atl gene for autolysin, complete cds and other ORFs	100	188
421	2	1262	2509	[gb U43098]	Transposon Tn5404 and insertion sequences IS1181 and IS1182 (from Staphylococcus aureus) DNA	99	1248
422	1	2	325	[gb K02985]	S aureus (strain RN450) transposon Tn554 insertion site	96	200
427	1	865	434	[dbj C28879 STAP]	Staphylococcus aureus gene for penicillin-binding protein 1, complete cds	100	432
427	2	1829	1122	[dbj C28879 STAP]	Staphylococcus aureus gene for penicillin-binding protein 1, complete cds	100	151
435	1	2	808	[dbj D86240 D862]	Staphylococcus aureus gene for unknown function and dit operon dltA, dltB, dltC and dltD genes, complete cds	100	556
435	2	832	999	[dbj D86240 D862]	Staphylococcus aureus gene for unknown function and dit operon dltA, dltB, dltC and dltD genes, complete cds	100	134
436	1	1341	685	[emb X17688 SAFE]	S aureus factor essential for expression of methicillin resistance (fema) gene, complete cds, and tnpA gene, 3' end	97	657

TABLE I

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	Percent ident	ORF nt length
436	2	2401	1657	[emb]N17688[SAPF]	S. aureus factor essential for expression of methicillin resistance (fema) gene, complete cds, and trpA gene, 3' and	100	294
442	1	347	1100	[emb]X72700[SAPV]	S. aureus genes for S and F components of Panton-Valentine leucocidins	84	204
445	2	1906	2178	[gb]LC1055]	Staphylococcus aureus gamma-hemolysin components A, B and C (hlyA, hlyB, hlyC) genes, complete cds	98	187
447	1	167	1078	[gb]U19770]	Staphylococcus aureus pyrrolidone carboxyl peptidase (pcp) gene, complete cds	100	514
447	2	1176	1784	[gb]U19770]	Staphylococcus aureus pyrrolidone carboxyl peptidase (pcp) gene, complete cds	96	597
454	3	7309	4319	[emb]218852[SACF]	S. aureus gene for clumping factor	75	653
472	4	7896	5479	[gb]L25288]	Staphylococcus aureus gyrase-like protein alpha and beta subunit (gria and grib) genes, complete cds	99	2418
472	5	8120	6792	[gb]L25288]	Staphylococcus aureus gyrase-like protein alpha and beta subunit (gria and grib) genes, complete cds	99	1328
475	2	566	889	[emb]X52543[SAG]	S. aureus agrA, agrB and hld genes	100	76
481	4	1922	1560	[emb]X64172[SARP]	S. aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	250
481	5	1244	1534	[emb]X64172[SARP]	S. aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA directed RNA polymerase beta & beta' chains	100	224
487	2	1388	1188	[gb]NR3994]	Staphylococcus aureus prolipoprotein signal peptidase (lsp) gene, complete cds	98	72
489	1	2737	1370	[gb]U21221]	Staphylococcus aureus hyaluronate lyase (hyla) gene, complete cds	99	1368
503	2	1135	653	[gb]NR3994]	Staphylococcus aureus prolipoprotein signal peptidase (lsp) gene, complete cds	100	108
511	3	1613	2242	[gb]L14017]	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	84	323
511	4	3122	2700	[gb]S76213]	esp23-alkaline shock protein 23 (methicillin resistant) (staphylococcus aureus, 912, Ganonic, 1360 nt)	96	423
520	2	758	1297	[emb]X72014[SAFI]	S. aureus fib gene for fibrinogen-binding protein	99	540
520	3	1436	1801	[emb]X72013[SAPI]	S. aureus fib gene for fibrinogen-binding protein	99	221
526	1	2150	1092	[dbj]D17366[STAA]	Staphylococcus aureus atl gene for autolysin, complete cds and other ORFs	99	641

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	Percent ident	ORF nt length	ORF nt length
528	2	58	963	gb U19100	Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank	99	260	906
528	3	1098	2870	gb U19100	Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank	99	866	1773
530	1	3	434	gb U19179	Staphylococcus aureus chorismate synthase (aroC) and nucleoside diphosphate kinase (ndk) genes, complete cds, dehydroquinase synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (garCC) genes, partial cds	99	432	432
530	2	1211	2395	gb U19179	Staphylococcus aureus chorismate synthase (aroC) and nucleoside diphosphate kinase (ndk) genes, complete cds, dehydroquinase synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (garCC) genes, partial cds	91	1185	1185
530	3	2409	2801	gb U19179	Staphylococcus aureus chorismate synthase (aroC) and nucleoside diphosphate kinase (ndk) genes, complete cds, dehydroquinase synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (garCC) genes, partial cds	88	181	393
530	4	2690	3484	gb U05004	Staphylococcus aureus dehydroquinase synthase (aroB) gene, 3' end cds; 3-phosphohikimate-1-carboxyvinyltransferase (aroA) gene, complete cds; ORF1, complete cds	100	75	795
530	5	3482	4792	gb U05004	Staphylococcus aureus dehydroquinase synthase (aroB) gene, 3' end cds; 3-phosphohikimate-1-carboxyvinyltransferase (aroA) gene, complete cds; ORF3, complete cds	99	905	1311
530	6	4790	5380	gb U05004	Staphylococcus aureus dehydroquinase synthase (aroB) gene, 3' end cds; 3-phosphohikimate-1-carboxyvinyltransferase (aroA) gene, complete cds; ORF3, complete cds	100	196	591
539	1	3	338	emb X76490 SAGL	S. aureus (bb270) glnA and glnR genes	99	316	316
539	2	336	527	emb X76490 SAGL	S. aureus (bb270) glnA and glnR genes	100	189	192
534	1	727	365	gb U73374	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	100	54	163
534	2	2175	1252	gb U73374	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	99	518	924
534	3	1574	1374	gb U73374	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	96	122	201
584	2	1019	705	gb U21221	Staphylococcus aureus hyaluronate lyase (hysA) gene, complete cds	99	306	315
587	3	1475	4286	emb 218852 SACF	S. aureus gene for clumping factor	98	2588	2814
598	3	3881	1953	dbj D24879 STAP	Staphylococcus aureus gene for penicillin-binding protein 1, complete cds	99	1873	1929

TABLE 1

S aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
605	1	2	745	[U01096240]D862	Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds	98	338	744
609	1	1628	816	[emb]X176490[SAGU]	S aureus (bb270) glnA and glnR genes	100	495	813
614	1	1280	642	[gi]M32103	Staphylococcus aureus lac repressor (lacR) gene, complete cds and lacA repressor (lacA), partial cds	99	639	639
626	1	2508	1255	[gb]H631761	Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds	100	225	1254
626	2	3115	2284	[gb]H63176	Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds	99	838	1012
629	1	1999	1001	[emb]X17688[SAFE]	S aureus factor essential for expression of methicillin resistance (femA) gene, complete cds, and trpA gene, 3' end	99	990	999
629	2	1607	1195	[emb]X17688[SAFE]	S aureus factor essential for expression of methicillin resistance (femA) gene, complete cds, and trpA gene, 3' end	98	194	213
631	2	5126	3228	[emb]Z18852[SACF]	S aureus gene for clumping factor	82	489	1899
632	1	3	551	[emb]Z10588[SAST]	S aureus (RM4220) genes for potential ABC transporter and potential membrane spanning protein	99	549	549
634	2	529	1323	[emb]Z10588[SAST]	S aureus (RM4220) genes for potential ABC transporter and potential membrane spanning protein	99	795	795
651	1	1909	1070	[gb]L191001	Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank	99	478	840
657	2	1800	1105	[gb]L140171	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	84	456	696
662	1	908	456	[emb]X13404[SAHL]	Staphylococcus aureus hib gene for beta-hemolysin	100	369	453
662	2	210	475	[emb]X13404[SAHL]	Staphylococcus aureus hib gene for beta-hemolysin	100	246	246
662	3	746	1399	[emb]X13404[SAHL]	Staphylococcus aureus hib gene for beta-hemolysin	99	653	654
682	1	956	480	[gb]H631771	S aureus sigma factor (pIac) gene, complete cds	100	116	477
685	1	1182	592	[gi]U650001	Staphylococcus aureus type-I signal peptidase SpsA (spsA) gene, end type-I signal peptidase SpsB (spsB) gene, complete cds	98	534	591
685	2	1716	1153	[gi]U650001	Staphylococcus aureus type-I signal peptidase SpsA (spsA) gene, end type-I signal peptidase SpsB (spsB) gene, complete cds	96	564	564
697	1	3	527	[gb]H631771	S aureus sigma factor (pIac) gene, complete cds	100	195	525
697	2	485	784	[gb]H631771	S aureus sigma factor (pIac) gene, complete cds	97	280	300

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	Percent ident	ORF nt length
710	1	15	503	[dbj D86240 D862	Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds	99	489
733	1	26	205	[gb H60252	Staphylococcus aureus norA199 gene (which mediates active efflux of fluorquinolones), complete cds	97	180
741	1	1736	1197	[cbj U83951 STAL	Staphylococcus aureus DNA for LukM component, LukP-PV like component, complete cds	81	522
752	1	1	636	[emb Y00356 SASP	Staphylococcus aureus V8 serine protease gene	99	618
752	2	548	956	[emb Y00356 SASP	Staphylococcus aureus V8 serine protease gene	99	340
756	1	1308	709	[emb X01645 SATO	Staphylococcus aureus (Wood 46) gene for alpha-toxin	98	600
777	1	1582	950	[emb Z49245 SA42	S. aureus partial sod gene for superoxide dismutase	99	633
780	1	1111	557	[gb U20503	Staphylococcus aureus MHC class II analog gene, complete cds	86	550
784	1	73	683	[gb U63529	Staphylococcus aureus novel antigen gene, complete cds	99	635
797	1	182	544	[dbj D14711 STAN	Staphylococcus aureus HSP10 and HSP60 genes	98	363
798	1	532	302	[emb X58434 SAPD	S. aureus pnhC, pnhC and pnhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	95	231
823	1	3	467	[gb S77055	tracF cluster: dnaA replisome assembly protein...gyrB-DNA gyrase beta subunit [Staphylococcus aureus, YB886, Genomic, 3573 nt]	99	465
848	1	348	175	[gb L25288	Staphylococcus aureus gyrase-like protein alpha and beta subunit (gria and griB) genes, complete cds	99	174
848	2	476	318	[gb L25288	Staphylococcus aureus gyrase-like protein alpha and beta subunit (gria and griB) genes, complete cds	100	159
866	1	792	397	[emb X64172 SARP	S. aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta L beta' chains	95	395
883	1	1	285	[dbj D90119 STAN	S. aureus norA gene	99	285
884	1	606	334	[emb X52543 SAG	S. aureus agrA, agrB and hid genes	98	273
884	2	716	522	[emb X52543 SAG	S. aureus agrA, agrB and hid genes	100	195
912	2	517	681	[emb Z30588 SAST	S. aureus (RM4220) genes for potential ABC transporter and potential membrane spanning protein	99	165
917	1	2	265	[gb M64724	S. aureus lagatose 6-phosphate isomerase gene, complete cds	99	264
917	2	238	396	[gb M64724	S. aureus lagatose 6-phosphate isomerase gene, complete cds	95	159
918	1	2426	1215	[emb X93305 SAPT	S. aureus plst and plstI genes	99	1212

TABLE I

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	Percent ident	ORF nt length
967	1	1	411	[dbj 090119 STAN S. aureus nora gene		97	195
991	1	672	337	[emb X52543 SAGG S. aureus agrA, agrB and hld genes		99	336
1000	1	1117	845	[gb U14017	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	78	190
1001	1	498	265	[dbj D86240 D862	Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds	99	234
1010	1	1	285	[gb U21221	Staphylococcus aureus hyaluronate lyase (hyaA) gene, complete cds	99	224
1046	1	656	330	[emb X72700 SAPV S. aureus genes for S and F components of Panton-Valentine leucocidins		85	205
1060	1	480	286	[emb X58434 SAPD	S. aureus pdhA, pdhC and pdhD genes for pyruvate decarboxylase, dihydroliipoamide acetyltransferase and dihydroliipoamide dehydrogenase	99	180
1073	1	1176	589	[gb X02985	S. aureus (strain RN450) transposon Tn534 insertion site	100	131
1079	1	3	230	[dbj D86240 D862	Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds	99	228
1079	2	218	484	[dbj D86240 D862	Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds	100	267
1079	1	460	645	[dbj D86240 D862	Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds	100	186
1092	1	289	146	[emb X58434 SAPD	S. aureus pdhA, pdhC and pdhD genes for pyruvate decarboxylase, dihydroliipoamide acetyltransferase and dihydroliipoamide dehydrogenase	98	124
1143	1	1	243	[gb H63177	S. aureus sigma factor (plac) gene, complete cds	99	243
1157	1	2	136	[emb X48003 SADN S. aureus gene for DNA polymerase III		97	127
1189	1	720	361	[gb S74031	[nora-Nora (ISP794)] Staphylococcus aureus, NCTC 8325, Insertion, 1820 nt	99	360
1190	1	2	283	[gb H21854	S. aureus agr gene encoding an accessory gene regulator protein, complete cds	100	282
1190	2	1127	888	[emb X52543 SAGG S. aureus agrA, agrB and hld genes		100	240
1225	1	2	163	[emb X17679 SACO Staphylococcus aureus coa gene for coagulase		97	124
1243	1	2	529	[dbj D86240 D862	Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds	99	495
1244	1	1	210	[gb S74031	[nora-Nora (ISP794)] Staphylococcus aureus, NCTC 8325, Insertion, 1820 nt	100	210
1301	1	41	472	[emb X76470 DAGL S. aureus (bb270), glpA and glpB genes		99	299

TABLE I

S. aureus - coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	Percent Ident	HSP nt length	ORF nt length
1315	1	1A	326	emb X64172 SARP	S aureus rplL,orf202, rpoB1(f) and rpoC genes for ribosomal protein L7/L12, hypothetical, protein ORF202, DNA-directed RNA polymerase beta L beta' chains	98	277	309
1519	1	2	175	dbj D28879 STAP	Staphylococcus aureus gene for penicillin-binding protein 1, complete cds	98	139	174
1663	1	1346	675	dbj D6240 D862	Staphylococcus aureus gene for unknown function and dit operon dltA, dltB, dltC and dltD genes, complete cds	98	672	672
1797	1	644	324	gb U73374	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	99	321	321
1857	1	1	192	gb U90536	Staphylococcus aureus alpha-hemolysin gene, 3' end	98	192	192
1923	1	2	181	emb X1768A SAFE	S aureus factor essential for expression of methicillin resistance (fema) gene, complete cds, and trpA gene, 3' end	100	180	180
1957	1	2	346	gb U60589	Staphylococcus aureus novel antigen gene, complete cds	99	345	345
1988	1	1	402	dbj D86240 D862	Staphylococcus aureus gene for unknown function and dit operon dltA, dltB, dltC and dltD genes, complete cds	100	402	402
2100	1	414	208	gb U63177	S aureus sigma factor (plac) gene, complete cds	99	207	207
2199	1	1	402	gb U66664	Staphylococcus aureus DNA fragment with class II promoter activity	99	131	402
2537	1	308	156	emb X1768A SAFE	S aureus factor essential for expression of methicillin resistance (fema) gene, complete cds, and trpA gene, 3' end	99	153	153
2891	1	2	400	gb U25426	Staphylococcus aureus penicillin-binding protein 2 (pbp2) gene, complete cds	99	399	399
2950	1	778	398	dbj D30690 STAN	Staphylococcus aureus genes for ORF37, ORF38, ORF39, ORF40, ORF41, complete cds	100	358	381
2971	1	3	398	gb U51132	Staphylococcus aureus o-succinylbenzoic acid CoA ligase (menc), and o-succinylbenzoic acid synthetase (menc) genes, complete cds	97	272	396
2978	1	618	328	gb U31979	Staphylococcus aureus chorismate synthase (aroc) and nucleoside diphosphate kinase (ndk) genes, complete cds, dehydroquinase synthase (arob) and geranylgeranyl pyrophosphate synthetase homolog (gerCC) genes, partial cds	98	250	291
2985	1	832	464	emb X17679 SACO	Staphylococcus aureus coa gene for coagulase	98	347	369
3006	1	2170	1784	gb U11779	Staphylococcus aureus methicillin-resistant ATCC 33952 clone RNNV30 16S-23S rRNA spacer region	87	82	387
3008	1	474	238	dbj D30690 STAN	Staphylococcus aureus genes for ORF37, ORF38, ORF39, ORF40, ORF41, complete cds	88	178	237
3008	2	451	281	dbj D30690 STAN	Staphylococcus aureus genes for ORF37, ORF38, ORF39, ORF40, ORF41, complete cds	97	120	171

TABLE I

S. aureus - coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	Percent ident	ORF nt length
3011	1	793	398	emb X62992 SAFN	S. aureus fnbB gene for fibronectin binding protein B	93	72
3019	1	2	235	gb J01479	S. aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	97	234
3023	1	81	233	gb U06451	Staphylococcus aureus proline permease homolog (putP) gene, complete cds	87	100
3029	1	90	287	gb U51133	Staphylococcus aureus phosphoenolpyruvate carboxylase (pckA) gene, complete cds	100	135
3039	1	18	164	gb U51133	Staphylococcus aureus phosphoenolpyruvate carboxylase (pckA) gene, complete cds	97	135
3039	2	70	327	gb U51133	Staphylococcus aureus phosphoenolpyruvate carboxylase (pckA) gene, complete cds	77	183
3056	1	3	215	emb X64172 SARP	S. aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	213
3059	1	1	261	dbj 030690 STAN	Staphylococcus aureus genes for ORF17, HSP10, HSP70, HSP40, ORF35, complete cds	98	234
3073	1	27	284	gb U06451	Staphylococcus aureus proline permease homolog (putP) gene, complete cds	99	229
3074	1	2	397	emb X64172 SARP	S. aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	96	250
3088	1	3	239	dbj D86727 D867	Staphylococcus aureus DNA for DNA polymerase III, complete cds	95	215
3077	1	444	244	emb Z18003 SAM	S. aureus gene for DNA polymerase III	97	160
3102	1	307	355	gb J01479	S. aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	97	142
3121	1	568	398	emb X58434 SAPD	S. aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	100	88
3125	1	463	233	emb X89233 SARP	S. aureus DNA for rpoC gene	98	192
3133	1	2	175	emb Z18852 SACF	S. aureus gene for clumping factor	96	154
3160	1	420	211	dbj 010489 STAC	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	89	197
3176	1	1	378	emb X58434 SAPD	S. aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	96	91
3192	1	420	211	gb J01479	S. aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	98	72
3210	1	3	143	gb J06714	Staphylococcus aureus peptidoglycan hydrolase gene, complete cds	96	141

TABLE I



S aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	percent ident	ORF nt length
3232	3	2106	1282	gb U14017	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	71	257
333A	1	2	394	emb X89333 SARP	S aureus DNA for rpoC gene	99	396
334J	1	392	634	gb U11510	Staphylococcus aureus transfer RNA sequence with two tRNAs	99	102
3355	1	637	320	emb Z18852 SACF	S aureus gene for clumping factor	99	307
3359	1	3	182	emb X17679 SACO	Staphylococcus aureus coa gene for coagulase	100	141
3359	2	95	313	emb X17679 SALO	Staphylococcus aureus coa gene for coagulase	98	174
3363	1	278	141	gb U35773	Staphylococcus aureus prolL protein diacylglycerol transferase (lgt) gene, complete cds	100	79
3363	2	527	363	gb U35773	Staphylococcus aureus prolipoprotein diacylglycerol transferase (lgt) gene, complete cds	98	162
3366	1	3	422	emb X16457 SAST	Staphylococcus aureus gene for staphylocoagulase	98	175
3388	1	2	262	gb U43098	Transposon Tn5404 and insertion sequences IS1181 and IS1182 (from Staphylococcus aureus) DNA	99	253
3393	1	3	150	emb U34479	S aureus enzyme Trf-lac (lacF), enzyme Trf-lac (lacF), and phospho-beta-galactosidase (lacG) genes, complete cds	99	145
3600	1	758	381	emb Z18852 SACF	S aureus gene for clumping factor	72	346
3602	1	788	396	emb Z18852 SACF	S aureus gene for clumping factor	98	319
3606	1	1013	528	emb Z18852 SACF	S aureus gene for clumping factor	84	403
3682	1	3	236	emb X64172 SARP	S aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	231
3682	2	224	415	emb X64172 SARP	S aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	112
369J	1	758	423	emb X62992 SAPN	S aureus fnbB gene for fibronectin binding protein B	100	229
3702	1	593	354	gb U11510	Staphylococcus aureus transfer RNA sequence with two tRNAs	94	81
3725	1	924	463	emb Z18852 SACF	S aureus gene for clumping factor	71	367
3761	1	809	450	gb U14017	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	85	333
3767	1	1	402	emb X64172 SARP	S aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	98	387

TABLE I

aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	Percent ident	USF nt length	ORF nt length
3775	1	2	286	emb X64172 SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	227	285
3786	1	456	229	dbj D10489 STAG	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	100	204	228
3786	2	512	366	dbj D10489 STAG	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	95	123	177
3798	1	3	251	emb X17679 SACO	Staphylococcus aureus coa gene for coagulase	99	249	249
3811	1	793	398	gb J04151	S.aureus fibronectin-binding protein (fnbA) mRNA, complete cds	98	396	396
3819	1	184	402	emb X68425 SA23	S.aureus gene for 23S rRNA	99	161	219
3844	1	932	468	gb U0826	Staphylococcus aureus elastin binding protein (ebpS) gene, complete cds	87	204	465
3845	1	1	381	emb X58434 SARD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	94	356	381
3856	1	798	400	gb U14017	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	76	192	399
3859	1	1049	573	emb Z18852 SACF	S.aureus gene for clumping factor	85	387	477
3871	1	650	327	gb U76714	Staphylococcus aureus peptidoglycan hydrolase gene, complete cds	100	299	324
3876	1	2	253	dbj D10489 STAG	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	100	217	252
3877	1	572	268	gb U03479	S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	97	209	285
3878	1	1	237	emb X58434 SARD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	96	155	237
3888	1	3	173	emb X16457 SAST	Staphylococcus aureus gene for staphylocoagulase	98	171	171
3893	1	1	183	emb X69233 SARP	S.aureus DNA for rpoC gene	100	170	183
3893	2	181	357	emb X69233 SARP	S.aureus DNA for rpoC gene	98	79	177
3894	1	3	485	emb X64172 SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	450	483
3905	1	416	420	gb U04151	S.aureus fibronectin-binding protein (fnbA) mRNA, complete cds	99	411	417
3905	1	48	220	gb U05004	Staphylococcus aureus dehydroquinase synthase (aroB) gene, 3'-phosphohikimate-1-carboxyvinyltransferase (aroA) gene, complete cds; ORF3, complete cds	100	159	192
3905	2	188	400	gb U05004	Staphylococcus aureus dehydroquinase synthase (aroB) gene, 3'-phosphohikimate-1-carboxyvinyltransferase (aroA) gene, complete cds; ORF3, complete cds	97	88	213

TABLE I

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	Percent ident	HSP nt length	ORF nt length
3910	1	3	359	emb[X58434]SAPD	S aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydroliipoamide acetyltransferase and dihydroliipoamide dehydrogenase	99	278	357
3915	1	1	310	gb[L14017]	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	75	175	330
3964	1	691	347	emb[Z48003]SADN	S aureus gene for DNA polymerase III	100	295	345
4007	1	199	390	emb[X16457]SAST	Staphylococcus aureus gene for staphylocoagulase	98	163	392
4036	1	1	371	dbj[D10489]STAG	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	99	339	369
4046	1	692	348	emb[Z18852]SACF	S aureus gene for clumping factor	87	221	345
4060	1	1	375	emb[Z18852]SACF	S aureus gene for clumping factor	96	271	375
4061	1	860	432	emb[Z48003]SADN	S aureus gene for DNA polymerase III	99	429	429
4082	1	606	304	gb[L14017]	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	75	198	303
4085	1	58	402	gb[U11786]	Staphylococcus aureus methicillin-resistance protein (mecR) gene and tRNA spacer region	98	127	345
4088	1	2	301	gb[J143098]	Transposon Tn5404 and insertion sequences IS1181 and IS1182 (from Staphylococcus aureus) DNA	99	227	300
4093	1	2	277	emb[X58434]SAPD	S aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydroliipoamide acetyltransferase and dihydroliipoamide dehydrogenase	99	276	276
4097	1	1	402	emb[Z18852]SACF	S aureus gene for clumping factor	74	307	402
4116	1	22	402	gi[L05004]	Staphylococcus aureus dehydroquinase synthase (aroH) gene, 3' end cds; 3-phosphoshikimate-1-carboxyvinyltransferase (aroA) gene, complete cds; ORF3, complete cds	98	157	381
4125	1	240	401	gb[U71374]	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, complete cds	100	86	162
4149	1	35	247	gb[J04151]	S aureus fibronectin-binding protein (fnbA) mRNA, complete cds	99	200	213
4151	1	629	366	gb[L14017]	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	87	150	264
4154	1	754	398	emb[X64172]SARP	S aureus rplL, rplM, rplN, rplO, rplP, rplQ, rplR, rplS, rplT, rplU, rplV, rplW, rplX, rplY, rplZ, rpl1, rpl2, rpl3, rpl4, rpl5, rpl6, rpl7, rpl8, rpl9, rpl10, rpl11, rpl12, rpl13, rpl14, rpl15, rpl16, rpl17, rpl18, rpl19, rpl20, rpl21, rpl22, rpl23, rpl24, rpl25, rpl26, rpl27, rpl28, rpl29, rpl30, rpl31, rpl32, rpl33, rpl34, rpl35, rpl36, rpl37, rpl38, rpl39, rpl40, rpl41, rpl42, rpl43, rpl44, rpl45, rpl46, rpl47, rpl48, rpl49, rpl50, rpl51, rpl52, rpl53, rpl54, rpl55, rpl56, rpl57, rpl58, rpl59, rpl60, rpl61, rpl62, rpl63, rpl64, rpl65, rpl66, rpl67, rpl68, rpl69, rpl70, rpl71, rpl72, rpl73, rpl74, rpl75, rpl76, rpl77, rpl78, rpl79, rpl80, rpl81, rpl82, rpl83, rpl84, rpl85, rpl86, rpl87, rpl88, rpl89, rpl90, rpl91, rpl92, rpl93, rpl94, rpl95, rpl96, rpl97, rpl98, rpl99, rpl100, rpl101, rpl102, rpl103, rpl104, rpl105, rpl106, rpl107, rpl108, rpl109, rpl110, rpl111, rpl112, rpl113, rpl114, rpl115, rpl116, rpl117, rpl118, rpl119, rpl120, rpl121, rpl122, rpl123, rpl124, rpl125, rpl126, rpl127, rpl128, rpl129, rpl130, rpl131, rpl132, rpl133, rpl134, rpl135, rpl136, rpl137, rpl138, rpl139, rpl140, rpl141, rpl142, rpl143, rpl144, rpl145, rpl146, rpl147, rpl148, rpl149, rpl150, rpl151, rpl152, rpl153, rpl154, rpl155, rpl156, rpl157, rpl158, rpl159, rpl160, rpl161, rpl162, rpl163, rpl164, rpl165, rpl166, rpl167, rpl168, rpl169, rpl170, rpl171, rpl172, rpl173, rpl174, rpl175, rpl176, rpl177, rpl178, rpl179, rpl180, rpl181, rpl182, rpl183, rpl184, rpl185, rpl186, rpl187, rpl188, rpl189, rpl190, rpl191, rpl192, rpl193, rpl194, rpl195, rpl196, rpl197, rpl198, rpl199, rpl200, rpl201, rpl202, rpl203, rpl204, rpl205, rpl206, rpl207, rpl208, rpl209, rpl210, rpl211, rpl212, rpl213, rpl214, rpl215, rpl216, rpl217, rpl218, rpl219, rpl220, rpl221, rpl222, rpl223, rpl224, rpl225, rpl226, rpl227, rpl228, rpl229, rpl230, rpl231, rpl232, rpl233, rpl234, rpl235, rpl236, rpl237, rpl238, rpl239, rpl240, rpl241, rpl242, rpl243, rpl244, rpl245, rpl246, rpl247, rpl248, rpl249, rpl250, rpl251, rpl252, rpl253, rpl254, rpl255, rpl256, rpl257, rpl258, rpl259, rpl260, rpl261, rpl262, rpl263, rpl264, rpl265, rpl266, rpl267, rpl268, rpl269, rpl270, rpl271, rpl272, rpl273, rpl274, rpl275, rpl276, rpl277, rpl278, rpl279, rpl280, rpl281, rpl282, rpl283, rpl284, rpl285, rpl286, rpl287, rpl288, rpl289, rpl290, rpl291, rpl292, rpl293, rpl294, rpl295, rpl296, rpl297, rpl298, rpl299, rpl300, rpl301, rpl302, rpl303, rpl304, rpl305, rpl306, rpl307, rpl308, rpl309, rpl310, rpl311, rpl312, rpl313, rpl314, rpl315, rpl316, rpl317, rpl318, rpl319, rpl320, rpl321, rpl322, rpl323, rpl324, rpl325, rpl326, rpl327, rpl328, rpl329, rpl330, rpl331, rpl332, rpl333, rpl334, rpl335, rpl336, rpl337, rpl338, rpl339, rpl340, rpl341, rpl342, rpl343, rpl344, rpl345, rpl346, rpl347, rpl348, rpl349, rpl350, rpl351, rpl352, rpl353, rpl354, rpl355, rpl356, rpl357, rpl358, rpl359, rpl360, rpl361, rpl362, rpl363, rpl364, rpl365, rpl366, rpl367, rpl368, rpl369, rpl370, rpl371, rpl372, rpl373, rpl374, rpl375, rpl376, rpl377, rpl378, rpl379, rpl380, rpl381, rpl382, rpl383, rpl384, rpl385, rpl386, rpl387, rpl388, rpl389, rpl390, rpl391, rpl392, rpl393, rpl394, rpl395, rpl396, rpl397, rpl398, rpl399, rpl400, rpl401, rpl402, rpl403, rpl404, rpl405, rpl406, rpl407, rpl408, rpl409, rpl410, rpl411, rpl412, rpl413, rpl414, rpl415, rpl416, rpl417, rpl418, rpl419, rpl420, rpl421, rpl422, rpl423, rpl424, rpl425, rpl426, rpl427, rpl428, rpl429, rpl430, rpl431, rpl432, rpl433, rpl434, rpl435, rpl436, rpl437, rpl438, rpl439, rpl440, rpl441, rpl442, rpl443, rpl444, rpl445, rpl446, rpl447, rpl448, rpl449, rpl450, rpl451, rpl452, rpl453, rpl454, rpl455, rpl456, rpl457, rpl458, rpl459, rpl460, rpl461, rpl462, rpl463, rpl464, rpl465, rpl466, rpl467, rpl468, rpl469, rpl470, rpl471, rpl472, rpl473, rpl474, rpl475, rpl476, rpl477, rpl478, rpl479, rpl480, rpl481, rpl482, rpl483, rpl484, rpl485, rpl486, rpl487, rpl488, rpl489, rpl490, rpl491, rpl492, rpl493, rpl494, rpl495, rpl496, rpl497, rpl498, rpl499, rpl500, rpl501, rpl502, rpl503, rpl504, rpl505, rpl506, rpl507, rpl508, rpl509, rpl510, rpl511, rpl512, rpl513, rpl514, rpl515, rpl516, rpl517, rpl518, rpl519, rpl520, rpl521, rpl522, rpl523, rpl524, rpl525, rpl526, rpl527, rpl528, rpl529, rpl530, rpl531, rpl532, rpl533, rpl534, rpl535, rpl536, rpl537, rpl538, rpl539, rpl540, rpl541, rpl542, rpl543, rpl544, rpl545, rpl546, rpl547, rpl548, rpl549, rpl550, rpl551, rpl552, rpl553, rpl554, rpl555, rpl556, rpl557, rpl558, rpl559, rpl560, rpl561, rpl562, rpl563, rpl564, rpl565, rpl566, rpl567, rpl568, rpl569, rpl570, rpl571, rpl572, rpl573, rpl574, rpl575, rpl576, rpl577, rpl578, rpl579, rpl580, rpl581, rpl582, rpl583, rpl584, rpl585, rpl586, rpl587, rpl588, rpl589, rpl590, rpl591, rpl592, rpl593, rpl594, rpl595, rpl596, rpl597, rpl598, rpl599, rpl600, rpl601, rpl602, rpl603, rpl604, rpl605, rpl606, rpl607, rpl608, rpl609, rpl610, rpl611, rpl612, rpl613, rpl614, rpl615, rpl616, rpl617, rpl618, rpl619, rpl620, rpl621, rpl622, rpl623, rpl624, rpl625, rpl626, rpl627, rpl628, rpl629, rpl630, rpl631, rpl632, rpl633, rpl634, rpl635, rpl636, rpl637, rpl638, rpl639, rpl640, rpl641, rpl642, rpl643, rpl644, rpl645, rpl646, rpl647, rpl648, rpl649, rpl650, rpl651, rpl652, rpl653, rpl654, rpl655, rpl656, rpl657, rpl658, rpl659, rpl660, rpl661, rpl662, rpl663, rpl664, rpl665, rpl666, rpl667, rpl668, rpl669, rpl670, rpl671, rpl672, rpl673, rpl674, rpl675, rpl676, rpl677, rpl678, rpl679, rpl680, rpl681, rpl682, rpl683, rpl684, rpl685, rpl686, rpl687, rpl688, rpl689, rpl690, rpl691, rpl692, rpl693, rpl694, rpl695, rpl696, rpl697, rpl698, rpl699, rpl700, rpl701, rpl702, rpl703, rpl704, rpl705, rpl706, rpl707, rpl708, rpl709, rpl710, rpl711, rpl712, rpl713, rpl714, rpl715, rpl716, rpl717, rpl718, rpl719, rpl720, rpl721, rpl722, rpl723, rpl724, rpl725, rpl726, rpl727, rpl728, rpl729, rpl730, rpl731, rpl732, rpl733, rpl734, rpl735, rpl736, rpl737, rpl738, rpl739, rpl740, rpl741, rpl742, rpl743, rpl744, rpl745, rpl746, rpl747, rpl748, rpl749, rpl750, rpl751, rpl752, rpl753, rpl754, rpl755, rpl756, rpl757, rpl758, rpl759, rpl760, rpl761, rpl762, rpl763, rpl764, rpl765, rpl766, rpl767, rpl768, rpl769, rpl770, rpl771, rpl772, rpl773, rpl774, rpl775, rpl776, rpl777, rpl778, rpl779, rpl780, rpl781, rpl782, rpl783, rpl784, rpl785, rpl786, rpl787, rpl788, rpl789, rpl790, rpl791, rpl792, rpl793, rpl794, rpl795, rpl796, rpl797, rpl798, rpl799, rpl800, rpl801, rpl802, rpl803, rpl804, rpl805, rpl806, rpl807, rpl808, rpl809, rpl810, rpl811, rpl812, rpl813, rpl814, rpl815, rpl816, rpl817, rpl818, rpl819, rpl820, rpl821, rpl822, rpl823, rpl824, rpl825, rpl826, rpl827, rpl828, rpl829, rpl830, rpl831, rpl832, rpl833, rpl834, rpl835, rpl836, rpl837, rpl838, rpl839, rpl840, rpl841, rpl842, rpl843, rpl844, rpl845, rpl846, rpl847, rpl848, rpl849, rpl850, rpl851, rpl852, rpl853, rpl854, rpl855, rpl856, rpl857, rpl858, rpl859, rpl860, rpl861, rpl862, rpl863, rpl864, rpl865, rpl866, rpl867, rpl868, rpl869, rpl870, rpl871, rpl872, rpl873, rpl874, rpl875, rpl876, rpl877, rpl878, rpl879, rpl880, rpl881, rpl882, rpl883, rpl884, rpl885, rpl886, rpl887, rpl888, rpl889, rpl890, rpl891, rpl892, rpl893, rpl894, rpl895, rpl896, rpl897, rpl898, rpl899, rpl900, rpl901, rpl902, rpl903, rpl904, rpl905, rpl906, rpl907, rpl908, rpl909, rpl910, rpl911, rpl912, rpl913, rpl914, rpl915, rpl916, rpl917, rpl918, rpl919, rpl920, rpl921, rpl922, rpl923, rpl924, rpl925, rpl926, rpl927, rpl928, rpl929, rpl930, rpl931, rpl932, rpl933, rpl934, rpl935, rpl936, rpl937, rpl938, rpl939, rpl940, rpl941, rpl942, rpl943, rpl944, rpl945, rpl946, rpl947, rpl948, rpl949, rpl950, rpl951, rpl952, rpl953, rpl954, rpl955, rpl956, rpl957, rpl958, rpl959, rpl960, rpl961, rpl962, rpl963, rpl964, rpl965, rpl966, rpl967, rpl968, rpl969, rpl970, rpl971, rpl972, rpl973, rpl974, rpl975, rpl976, rpl977, rpl978, rpl979, rpl980, rpl981, rpl982, rpl983, rpl984, rpl985, rpl986, rpl987, rpl988, rpl989, rpl990, rpl991, rpl992, rpl993, rpl994, rpl995, rpl996, rpl997, rpl998, rpl999, rpl1000, rpl1001, rpl1002, rpl1003, rpl1004, rpl1005, rpl1006, rpl1007, rpl1008, rpl1009, rpl1010, rpl1011, rpl1012, rpl1013, rpl1014, rpl1015, rpl1016, rpl1017, rpl1018, rpl1019, rpl1020, rpl1021, rpl1022, rpl1023, rpl1024, rpl1025, rpl1026, rpl1027, rpl1028, rpl1029, rpl1030, rpl1031, rpl1032, rpl1033, rpl1034, rpl1035, rpl1036, rpl1037, rpl1038, rpl1039, rpl1040, rpl1041, rpl1042, rpl1043, rpl1044, rpl1045, rpl1046, rpl1047, rpl1048, rpl1049, rpl1050, rpl1051, rpl1052, rpl1053, rpl1054, rpl1055, rpl1056, rpl1057, rpl1058, rpl1059, rpl1060, rpl1061, rpl1062, rpl1063, rpl1064, rpl1065, rpl1066, rpl1067, rpl1068, rpl1069, rpl1070, rpl1071, rpl1072, rpl1073, rpl1074, rpl1075, rpl1076, rpl1077, rpl1078, rpl1079, rpl1080, rpl1081, rpl1082, rpl1083, rpl1084, rpl1085, rpl1086, rpl1087, rpl1088, rpl1089, rpl1090, rpl1091, rpl1092, rpl1093, rpl1094, rpl1095, rpl1096, rpl1097, rpl1098, rpl1099, rpl1100, rpl1101, rpl1102, rpl1103, rpl1104, rpl1105, rpl1106, rpl1107, rpl1108, rpl1109, rpl1110, rpl1111, rpl1112, rpl1113, rpl1114, rpl1115, rpl1116, rpl1117, rpl1118, rpl1119, rpl1120, rpl1121, rpl1122, rpl1123, rpl1124, rpl1125, rpl1126, rpl1127, rpl1128, rpl1129, rpl1130, rpl1131, rpl1132, rpl1133, rpl1134, rpl1135, rpl1136, rpl1137, rpl1138, rpl1139, rpl1140, rpl1141, rpl1142, rpl1143, rpl1144, rpl1145, rpl1146, rpl1147, rpl1148, rpl1149, rpl1150, rpl1151, rpl1152, rpl1153, rpl1154, rpl1155, rpl1156, rpl1157, rpl1158, rpl1159, rpl1160, rpl1161, rpl1162, rpl1163, rpl1164, rpl1165, rpl1166, rpl1167, rpl1168, rpl1169, rpl1170, rpl1171, rpl1172, rpl1173, rpl1174, rpl1175, rpl1176, rpl1177, rpl1178, rpl1179, rpl1180, rpl1181, rpl1182, rpl1183, rpl1184, rpl1185, rpl1186, rpl1187, rpl1188, rpl1189, rpl1190, rpl1191, rpl1192, rpl1193, rpl1194, rpl1195, rpl1196, rpl1197, rpl1198, rpl1199, rpl1200, rpl1201, rpl1202, rpl1203, rpl1204, rpl1205, rpl1206, rpl1207, rpl1208, rpl1209, rpl1210, rpl1211, rpl1212, rpl1213, rpl1214, rpl1215, rpl1216, rpl1217, rpl1218, rpl1219, rpl1220, rpl1221, rpl1222, rpl1223, rpl1224, rpl1225, rpl1226, rpl1227, rpl1228, rpl1229, rpl1230, rpl1231, rpl1232, rpl1233, rpl1234, rpl1235, rpl1236, rpl1237, rpl1238, rpl1239, rpl1240, rpl1241, rpl1242, rpl1243, rpl1244, rpl1245, rpl1246, rpl1247, rpl1248, rpl1249, rpl1250, rpl1251, rpl1252, rpl1253, rpl1254, rpl1255, rpl1256, rpl1257, rpl1258, rpl1259, rpl1260, rpl1261, rpl1262, rpl1263, rpl1264, rpl1265, rpl1266, rpl1267, rpl1268, rpl1269, rpl1270, rpl1271, rpl1272, rpl1273, rpl1274, rpl1275, rpl1276, rpl1277, rpl1278, rpl1279, rpl1280, rpl1281, rpl1282, rpl1283, rpl1284, rpl1285, rpl1286, rpl1287, rpl1288, rpl1289, rpl1290, rpl1291, rpl1292, rpl1293, rpl1294, rpl1295, rpl1296, rpl1297, rpl1298, rpl1299, rpl1300, rpl1301, rpl1302, rpl1303, rpl1304, rpl1305, rpl1306, rpl1307, rpl1308, rpl1309, rpl1310, rpl1311, rpl1312, rpl1313, rpl1314, rpl1315, rpl1316, rpl1317, rpl1318, rpl1319, rpl1320, rpl1321, rpl1322, rpl1323, rpl1324, rpl1325, rpl1326, rpl1327, rpl1328, rpl1329, rpl1330, rpl1331, rpl1332, rpl1333, rpl1334, rpl1335, rpl1336, rpl1337, rpl1338, rpl1339, rpl1340, rpl1341, rpl1342, rpl1343, rpl1344, rpl1345, rpl1346, rpl1347, rpl1348, rpl1349, rpl1350, rpl1351, rpl1352, rpl1353, rpl1354, rpl1355, rpl1356, rpl1357, rpl1358, rpl1359, rpl1360, rpl1361, rpl1362, rpl1363, rpl1364, rpl1365, rpl1366, rpl1367, rpl1368, rpl1369, rpl1370, rpl1371, rpl1372, rpl1373, rpl1374, rpl1375, rpl1376, rpl1377, rpl1378, rpl1379, rpl1380, rpl1381, rpl1382, rpl1383, rpl1384, rpl1385, rpl1386, rpl1387, rpl1388, rpl1389, rpl1390, rpl1391, rpl1392, rpl1393, rpl1394, rpl1395, rpl1396, rpl1397, rpl1398, rpl1399, rpl1400, rpl1401, rpl1402, rpl1403, rpl1404, rpl1405, rpl1406, rpl1407, rpl1408, rpl1409, rpl1410, rpl1411, rpl1412, rpl1413, rpl1414, rpl1415, rpl1416, rpl1417, rpl1418, rpl1419, rpl1420, rpl1421, rpl1422, rpl1423, rpl1424, rpl1425, rpl1426, rpl1427, rpl1428, rpl1429, rpl1430, rpl1431, rpl1432, rpl1433, rpl1434, rpl1435, rpl1436, rpl1437, rpl1438, rpl1439, rpl1440, rpl1441, rpl1442, rpl1443, rpl1444, rpl1445, rpl1446, rpl1447, rpl1448, rpl1449, rpl1450, rpl1451, rpl1452, rpl1453, rpl1454, rpl1455, rpl1456, rpl1457, rpl1458, rpl1459, rpl1460, rpl1461, rpl1462, rpl1463, rpl1464, rpl1465, rpl1466, rpl1467, rpl1468, rpl1469, rpl1470, rpl1471, rpl1472, rpl1473, rpl1474, rpl1475, rpl1476, rpl1477, rpl1478, rpl1479, rpl1480, rpl1481, rpl1482, rpl1483, rpl1484, rpl1485, rpl1486, rpl1487, rpl1488, rpl1489, rpl1490, rpl1491, rpl1492, rpl1493, rpl1494, rpl1495, rpl1496, rpl1497, rpl1498, rpl1499, rpl1500, rpl1501, rpl1502, rpl1503, rpl1504, rpl1505, rpl1506, rpl1507, rpl1508, rpl1509, rpl1510, rpl1511, rpl1512, rpl1513, rpl1514, rpl1515, rpl1516, rpl1517, rpl1518, rpl1519, rpl1520, rpl1521, rpl1522, rpl1523, rpl1524, rpl1525, rpl1526, rpl1527, rpl1528, rpl1529, rpl1530, rpl1531, rpl1532, rpl1533, rpl1534, rpl1535, rpl1536, rpl1537, rpl1538, rpl1539, rpl1540, rpl1541, rpl1542, rpl1543, rpl1544, rpl1545, rpl1546, rpl1547, rpl1548, rpl1549, rpl1550, rpl1551, rpl1552, rpl1553, rpl1554, rpl1555, rpl1556, rpl1557, rpl1558, rpl1559, rpl1560, rpl1561, rpl1562, rpl1563, rpl1564, rpl1565, rpl1566, rpl1567, rpl1568, rpl1569, rpl1570, rpl1571, rpl1572, rpl1573, rpl1574, rpl1575, rpl1576, rpl1577, rpl1578, rpl1579, rpl1580, rpl1581, rpl1582, rpl1583, rpl1584, rpl1585, rpl1586, rpl1587, rpl1588, rpl1589, rpl1590, rpl1591, rpl1592, rpl1593, rpl1594, rpl1595, rpl1596, rpl1597, rpl1598, rpl1599, rpl1600, rpl1601, rpl1602, rpl1603, rpl1604, rpl1605, rpl1606, rpl1607, rpl1608, rpl1609, rpl1610, rpl1611, rpl1612, rpl1613, rpl1614, rpl1615, rpl1616, rpl1617, rpl1618, rpl1619, rpl1620, rpl1621, rpl1622, rpl1623, rpl1624, rpl1625, rpl1626, rpl1627, rpl1628, rpl1629, rpl1630, rpl1631, rpl1			

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	Percent ident	HSP nt length	ORF nt length
4203	1	1	255	emb X89233 SARP	S. aureus DNA for rpoC gene	99	239	255
4206	1	1	303	emb Z18852 SACF	S. aureus gene for clumping factor	100	276	303
4206	2	195	344	emb Z18852 SACF	S. aureus gene for clumping factor	95	65	150
4208	1	108	314	emb X56434 SAPD	S. aureus pdhB, pdHC and pdHD genes for pyruvate decarboxylase, dihydroliipoamide acetyltransferase and dihydroliipoamide dehydrogenase	89	76	207
4216	1	656	330	emb X56434 SAPD	S. aureus pdhB, pdHC and pdHD genes for pyruvate decarboxylase, dihydroliipoamide acetyltransferase and dihydroliipoamide dehydrogenase	98	126	327
4226	1	594	298	gb U11530	Staphylococcus aureus transfer RNA sequence with two tRNAs	97	132	297
4260	1	216	383	gb U11784	Staphylococcus aureus methicillin-resistant ATCC 33952 clone RN410 165-235 rRNA spacer region	83	141	168
4272	1	325	179	emb X24R003 SAUN	S. aureus gene for DNA polymerase III	100	164	177
4276	1	4	177	emb X16457 SAST	Staphylococcus aureus gene for staphylocoagulase	99	150	174
4277	1	1	270	emb X64172 SARP	S. aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	265	270
4282	1	691	377	emb X64172 SARP	S. aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	98	282	315
4291	1	379	191	emb X64172 SARP	S. aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	183	189
4295	1	3	329	emb X16457 SAST	Staphylococcus aureus gene for staphylocoagulase	94	144	327
4313	1	435	280	gb U11530	Staphylococcus aureus transfer RNA sequence with two tRNAs	100	94	156
4315	1	3	185	gb J03479	S. aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	100	158	183
4315	2	101	310	gb J03479	S. aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	98	75	210
4327	1	1	294	gb U43098	Transposon Tn5404 and insertion sequences IS1181 and IS1182 (from Staphylococcus aureus) DNA	98	294	294
4360	1	403	319	gb U02910	Staphylococcus aureus ATCC 25923 16S rRNA gene, partial sequence	100	116	285
4364	1	3	146	emb X64172 SARP	S. aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	95	140	144
4388	1	167	310	emb X62952 SAFN	S. aureus fnbB gene for fibronectin binding protein B	73	119	144

TABLE 1

S. aureus - coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	percent ident	HSP nt length	ORF nt length
4401	1	2	313	emb X62992 SAFN	S. aureus fnbB gene for fibronectin binding protein B	97	243	312
4421	1	36	281	dbj D12572 STA2	Staphylococcus aureus rna gene for 23S ribosomal RNA	100	112	246
4426	1	3	293	emb Z18852 SACF	S. aureus gene for clumping factor	85	185	291
4428	1	493	248	emb X64172 SARP	S. aureus rplL, rplL2, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	139	246
4462	1	2	271	emb X64172 SARP	S. aureus rplL, rplL2, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	270	270
4466	1	1	240	emb Z18852 SACF	S. aureus gene for clumping factor	99	231	210
4469	1	1	312	gb J03479	S. aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	99	265	312
4485	1	3	263	gb L41098	Transposon Tn5404 and insertion sequences IS1181 and IS1182 (from Staphylococcus aureus) DNA	98	259	261
4492	1	74	400	gb X86227	Staphylococcus aureus DNA gyrase B subunit (gyrB) RecF homologue (recF) and DNA gyrase A subunit (gyrA) gene, complete cds	85	104	127
4497	1	515	269	emb Z18852 SACF	S. aureus gene for clumping factor	99	213	267
4529	1	2	172	emb X64172 SARP	S. aureus rplL, rplL2, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	151	171
4547	1	1	300	emb X62992 SAFN	S. aureus fnbB gene for fibronectin binding protein B	100	157	300
4554	1	318	160	emb Z18852 SACF	S. aureus gene for clumping factor	84	126	159
4565	1	9	227	emb Z18852 SACF	S. aureus gene for clumping factor	84	213	219
4569	1	79	222	emb Z18852 SACF	S. aureus gene for clumping factor	98	127	144
4608	1	22	216	emb X58434 SAPD	S. aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	92	168	195
4614	1	464	204	emb Z18852 SACF	S. aureus gene for clumping factor	86	169	231
4623	1	105	302	gb J04151	S. aureus fibronectin-binding protein (fnbA) mRNA, complete cds	99	152	198
4632	1	18	206	gb J03479	S. aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	98	183	189
4656	1	1	222	emb Z18852 SACF	S. aureus gene for clumping factor	84	100	222
4687	1	2	166	gb J04151	S. aureus fibronectin-binding protein (fnbA) mRNA, complete cds	98	156	165

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	percent ident	HSP nt length	ORF nt length
4695	1	313	158	gk1114017	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	75	155	156
4703	1	1	153	emb X58434 SAPP	S. aureus pdhA, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	98	103	153

TABLE I

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
20	6	5089	4679	gi1511819	ORF1 [Staphylococcus bacteriophage phi 11]	100	100	411
149	3	2012	1577	gi1849703 B497	Int gene activator RlnA - bacteriophage phi 11	100	100	436
149	5	2109	1912	gi1166161	Bacteriophage phi-11 int gene activator [Staphylococcus acteriophage phi 11]	100	100	198
349	2	558	409	gi1166159	Integrase (int) [Staphylococcus bacteriophage phi 11]	100	100	150
398	1	1372	707	gi1166159	Integrase (int) [Staphylococcus bacteriophage phi 11]	100	99	666
398	2	783	1001	gi1455128	Excisionase (xis) [Staphylococcus bacteriophage phi 11]	100	100	219
502	4	1914	1744	gi1204912	H. influenzae predicted coding region H10640 (Haemophilus influenzae)	100	71	171
849	1	2	262	gi1373002	polyprotein [Bean common mosaic virus]	100	46	261
1149	1	277	140	gi1143359	protein synthesis initiation factor 2 (infB) [Bacillus subtilis] gi149319	100	62	138
2880	1	21	308	gi1862933	protein kinase C inhibitor-1 (Homo sapiens)	100	98	288
3085	1	428	216	gi1354211	PET112-like protein [Bacillus subtilis]	100	100	213
4168	2	571	398	gi1354211	PET112-like protein [Bacillus subtilis]	100	100	174
331	1	2	247	gi1426873	nusG gene product [Staphylococcus carnosus]	98	95	246
207	2	1272	1463	gi1460259	enolase [Bacillus subtilis]	97	90	192
331	2	392	850	gi1581618	hil protein [Staphylococcus carnosus]	97	93	456
366	1	39	215	gi166161	Bacteriophage phi-11 int gene activator [Staphylococcus acteriophage phi 11]	97	95	177
680	3	718	936	gi1426873	nusG gene product [Staphylococcus carnosus]	97	97	219
3578	1	284	144	gi1319250	large subunit of NAMF-dependent glutamate synthase [Plectononema boryanum]	97	79	141
157	1	321	518	gi1022726	unknown [Staphylococcus haemolyticus]	96	88	198
205	33	16470	16147	gi1165302	S10 [Bacillus subtilis]	96	91	324
3919	1	48	401	gi1871784	Clp-like ATP-dependent protease binding subunit [Bos taurus]	96	81	354
4133	1	810	417	gi1022726	unknown [Staphylococcus haemolyticus]	96	84	414
4168	1	708	355	gi1354211	PET112-like protein [Bacillus subtilis]	96	95	354
4207	1	312	157	gi1602031	similar to trimethylamine DH [Mycoplasma capricolum] pir184950184950 probable trimethylamine dehydrogenase (EC 5.99.7) - Mycoplasma capricolum (50C3) (fragment)	96	86	156

TABLE 2

S. aureus - putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
4227	2	152	331	gi1871784	Clp-like ATP-dependent protease binding subunit (Bos taurus)	96	91	180
4416	1	570	286	gi1022726	unknown (Staphylococcus haemolyticus)	96	84	285
22	1	858	430	gi1511070	UreG (Staphylococcus xyloosus)	95	88	429
22	7	4382	4036	gi1581787	lucase gamma subunit (Staphylococcus xyloosus)	95	79	327
82	6	8794	9114	gi1500081000	ribosomal protein S7 - Bacillus stearothermophilus	95	83	321
154	9	9280	7838	gi11354211	PEP12-like protein (Bacillus subtilis)	95	92	1443
186	3	2798	2055	gi1514656	serine O-acetyltransferase (Staphylococcus xyloosus)	95	87	744
205	5	4406	4014	gi142462	ribosomal protein S11 (Bacillus subtilis)	95	85	393
205	7	5017	6793	gi1142459	initiation factor 1 (Bacillus subtilis)	95	84	225
205	21	11365	10991	gi11044974	ribosomal protein L14 (Bacillus subtilis)	95	93	375
259	5	7288	6644	sp1479951YSEA_	HYPOTHETICAL PROTEIN IN SECA 5'-REGION (ORF1) (FRAGMENT)	95	85	645
302	3	795	1097	gi140186	homologous to E. coli ribosomal protein L27 (Bacillus subtilis) L143592 L27 ribosomal protein (Bacillus subtilis) tr[C21895]C21895 ribosomal protein L27 - Bacillus subtilis U P05657 BL27_BACSU 50S RIBOSOMAL PROTEIN L27 (BL30) (BL24). L140175 L24 gene prod	95	89	303
310	1	579	1523	gi11177684	chorismate mutase (Staphylococcus xyloosus)	95	92	945
414	1	2	163	gi148396[C483]	ribosomal protein L34 - Bacillus stearothermophilus	95	90	162
4185	2	125	277	gi11278841	glutamate synthase (GOGAT) (Porphyra purpurea)	95	86	153
22	2	1028	723	gi1511069	UreF (Staphylococcus xyloosus)	94	91	304
22	5	5046	3310	gi1410516	lucase alpha subunit (Staphylococcus xyloosus)	94	85	1737
60	4	815	1372	gi1666116	glucose kinase (Staphylococcus xyloosus)	94	87	558
205	18	10012	9536	gi11044978	ribosomal protein S8 (Bacillus subtilis)	94	78	477
326	4	3378	2542	gi1557492	dihydroxynaphthoic acid (DHNA) synthetase (Bacillus subtilis) gi143186	94	85	837
414	3	737	955	gi1467386	thiolphen and fucan oxidation (Bacillus subtilis)	94	77	219
426	3	2260	1823	gi1263908	putative (Staphylococcus epidermidis)	94	87	438
534	1	2	355	gi1633650	enzyme II (mannitol) (Staphylococcus carnosus)	94	84	354
1017	1	2	229	gi1149435	putative (Lactococcus lactis)	94	73	228
3098	1	330	184	gi1413952	lpa-28d gene product (Bacillus subtilis)	94	50	147

TABLE 2



S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
3232	1	630	316	gi1022725	unknown [Staphylococcus haemolyticus]	94	84	315
42	5	2089	2259	pir184839618483	ribosomal protein L33 - Bacillus stearothermophilus	93	81	171
101	2	1745	1383	gi155345	arsenic efflux pump protein [Plasmid pSK267]	93	82	363
205	24	12227	11865	sp14577RL16	50S RIBOSOMAL PROTEIN L16	93	83	363
259	4	8291	5673	gi1499335	secA protein [Staphylococcus carnosus]	93	85	2619
275	1	2226	1114	gi1433650	enzyme II (mammalian) [Staphylococcus carnosus]	93	86	1113
444	6	6207	5773	gi1022726	unknown [Staphylococcus haemolyticus]	93	81	435
481	1	152	622	gi146912	ribosomal protein L13 [Staphylococcus carnosus]	93	88	471
607	6	1674	2033	gi1022726	unknown [Staphylococcus haemolyticus]	93	81	360
653	1	973	488	gi1580890	translation initiation factor IF3 (AA 1-172) [Bacillus stearothermophilus]	93	77	486
1864	1	3	194	gi1306553	ribosomal protein aa1: subunit (Homo sapiens)	93	93	192
2997	1	28	300	gi1143390	[carbamyl] phosphate synthetase [Bacillus subtilis]	93	82	273
3232	2	907	586	gi1022725	unknown [Staphylococcus haemolyticus]	93	84	312
3761	2	794	621	gi1022725	unknown [Staphylococcus haemolyticus]	93	88	174
16	1	3	374	gi1142781	putative cytoplasmic protein: putative [Bacillus subtilis]	92	83	372
31	7	5915	6124	gi1136430	sp137954UVRA_BACSU EXCINUCLEASE ABC SUBUNIT B (DNA PROTEIN) FRAGMENT	92	46	210
56	119	26483	27381	gi1467401	unknown [Bacillus subtilis]	92	80	909
69	6	5882	6130	gi1530200	[crophoblastin [Ovis aries]	92	53	249
145	3	2568	2038	gi1022725	unknown [Staphylococcus haemolyticus]	92	80	531
171	3	2760	2362	gi1517475	D-amino acid transaminase [Staphylococcus haemolyticus]	92	86	399
205	12	7495	6962	gi149189	secY gene product [Staphylococcus carnosus]	92	85	534
205	19	10812	10255	gi1044976	ribosomal protein L5 [Bacillus subtilis]	92	82	558
219	1	710	357	gi1303812	Vqav [Bacillus subtilis]	92	88	354
344	3	1575	1805	gi1405474	CspC protein [Bacillus cereus]	92	85	231
699	1	20	161	gi1413999	ipa-75d gene product [Bacillus subtilis]	92	81	342
1343	1	2	160	pir145434A454	ribosomal protein L19 - Bacillus stearothermophilus	92	84	159

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
1958	1	524	264	gi1407908	ElfAcr [Staphylococcus xylosum]	92	80	261
3578	2	718	386	gi11339950	large subunit of NADH-dependent glutamate synthase [Plectonassa boryanum]	92	78	333
3585	1	644	324	gi11339950	large subunit of NADH-dependent glutamate synthase [Plectonassa boryanum]	92	83	321
3640	1	4	402	gi11022726	unknown [Staphylococcus haemolyticus]	92	81	399
4362	1	14	178	gi1450688	hadM gene of Ecoprr1 gene product [Escherichia coli] pir[S18437]S18437 hadM protein - Escherichia coli pir[S09629]S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	92	76	165
4446	1	358	182	gi11022725	unknown [Staphylococcus haemolyticus]	92	82	177
4549	1	462	232	gi11022726	unknown [Staphylococcus haemolyticus]	92	80	231
4626	1	3	224	gi11022725	unknown [Staphylococcus haemolyticus]	92	84	222
2	4	1980	4531	gi1535349	CodW [Bacillus subtilis]	91	74	552
28	3	2	1126	gi11001376	hypothetical protein [Synecocystis sp.]	91	78	1125
60	5	1354	1701	gi1122604	orf2 downstream of glucose kinase [Staphylococcus xylosum]	91	80	348
101	1	1989	1036	gi1150728	arsenic efflux pump protein [Plasmid p1258]	91	80	954
187	2	412	1194	gi1142559	ATP synthase alpha subunit [Bacillus megaterium]	91	79	783
205	22	11579	11298	gi140149	S17 protein (AA 1-87) [Bacillus subtilis]	91	83	282
206	7	8184	10262	gi11072418	glcA gene product [Staphylococcus carnosus]	91	83	2079
306	2	3885	2326	gi1143012	UMP synthetase [Bacillus subtilis]	91	78	1560
306	3	5319	3826	gi1467399	UMP dehydrogenase [Bacillus subtilis]	91	79	1494
310	3	2194	3207	gi1117685	ccpA gene product [Staphylococcus xylosum]	91	81	1014
343	4	2974	3150	gi1949574	sucrose repressor [Staphylococcus xylosum]	91	82	177
480	3	1606	3042	gi1433991	ATP synthase subunit beta [Bacillus subtilis]	91	85	1437
536	3	2026	1280	gi1143366	adenylosuccinate lyase (Pur-A) [Bacillus subtilis] pir[C29326]W28SDS adenylosuccinate lyase (EC 4.3.2.2) - Bacillus subtilis	91	79	747
552	1	1064	615	gi1297874	fructose-bisphosphate aldolase [Staphylococcus carnosus] pir[A4993]A49943 fructose-bisphosphate aldolase (EC 4.1.2.1) - Staphylococcus carnosus [strain TH300]	91	79	450
637	1	1	1536	gi1143597	CTP synthetase [Bacillus subtilis]	91	79	1536
859	1	21	359	gi1385178	unknown [Bacillus subtilis]	91	66	339

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
1327	1	339	530	gi 496558	orfX [Bacillus subtilis]	91	71	192
2515	1	466	275	gi 511070	ureG [Staphylococcus xylosus]	91	85	192
2594	1	2	202	gi 146824	beta-cystathionase [Escherichia coli]	91	75	201
3764	1	847	425	gi 1022725	unknown [Staphylococcus haemolyticus]	91	78	423
4011	1	127	495	gi 1022726	unknown [Staphylococcus haemolyticus]	91	79	369
4227	1	1	177	gi 296464	ATPase [Lactococcus lactis]	91	66	177
42	3	815	1033	gi 520401	catalase [Haemophilus influenzae]	90	86	219
51	8	3717	4607	gi 580899	OppP gene product [Bacillus subtilis]	90	74	891
129	3	5317	4001	gi 1146206	glutamate dehydrogenase [Bacillus subtilis]	90	76	1317
164	17	16628	16933	gi P05766 RS15_	30S RIBOSOMAL PROTEIN S15 (RS18)	90	74	306
171	5	2983	2819	gi 517475	D-amino acid transaminase [Staphylococcus haemolyticus]	90	78	165
205	4	4497	3550	gi 142463	RNA polymerase alpha-core-subunit [Bacillus subtilis]	90	76	948
205	6	4748	4410	gi 1044989	ribosomal protein S13 [Bacillus subtilis]	90	73	339
205	10	755	6406	gi 49189	IsecY gene product [Staphylococcus carnosus]	90	61	762
205	11	6645	6472	gi 49189	IsecY gene product [Staphylococcus carnosus]	90	78	174
205	27	11674	13345	gi 796157	Ribosomal Protein S19 [Bacillus subtilis]	90	79	348
205	31	15858	15496	gi 1165303	IS [Bacillus subtilis]	90	79	363
260	5	7023	5773	gi 1161380	IcaA [Staphylococcus epidermidis]	90	78	1251
299	6	3378	3947	gi 467440	phosphoribosylpyrophosphate synthetase [Bacillus subtilis] gi 40218 PRPP synthetase (AA 1-317) [Bacillus subtilis]	90	78	570
320	2	1025	1717	gi 312443	carbamoyl-phosphate synthase (glutamine-hydrolysing) [Bacillus aldolyticus]	90	75	693
330	4	1581	1769	gi 984963	beta-tubulin [Sporidiobolus pararoseus]	90	80	189
369	1	954	523	gi 53762 S347	L-serine dehydratase beta chain - Clostridium sp.	90	77	432
557	1	3	188	gi 1511589	M. jamaeensis predicted coding region M1624 [Methanococcus jamaeensis]	90	54	186
663	2	667	1200	gi 143786	cryptophanyl-tRNA synthetase (EC 6.1.1.2) [Bacillus subtilis] pi JTC481 YWS	90	73	534
717	1	1	261	gi 143065	hubat [Bacillus stearothermophilus]	90	79	261
745	4	1059	865	gi 1205433	H. influenzae predicted coding region H1190 [Haemophilus influenzae]	90	81	195

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
1007	1	386	565	gi1143366	adenylosuccinate lyase (PUB-8) [Bacillus subtilis] pf1C29J26 W2B5DS adenylsuccinate lyase (EC 4.3.2.1) - Bacillus subtilis	90	77	180
1034	1	579	331	gi11033122	ORF_729 [Escherichia coli]	90	50	249
1156	1	117	707	gi11477776	ClpP [Bacillus subtilis]	90	80	591
1180	1	408	205	gi11377831	unknown [Bacillus subtilis]	90	74	204
1251	1	1	462	gi1140046	phosphoglucose isomerase A (AA 1-449) [Bacillus stearothermophilus] trIS15936 NUB5A glucose-6-phosphate isomerase (EC 5.3.1.9) A - cillus stearothermophilus	90	75	462
2951	1	3	269	gi1144816	formyltetrahydrofolate synthetase (FTHS) (ttg start codon) (EC 3.4.3) [Mortella thermocellae]	90	76	267
3140	1	327	166	gi11070014	protein-dependent [Bacillus subtilis]	90	52	162
4594	1	3	233	gi11871784	Clp-like ATP-dependent protease binding subunit [Bos taurus]	90	76	231
87	1	1028	1750	gi11467327	unknown [Bacillus subtilis]	89	75	723
112	1	2	505	gi1151741	ATP-binding protein [Streptococcus mutans]	89	77	504
114	1	120	398	gi11303804	YqeQ [Bacillus subtilis]	89	75	279
128	4	3545	3757	gi11460257	fructose phosphate isomerase [Bacillus subtilis]	89	84	213
164	12	11667	12755	gi1139954	IF2 (aa 1-741) [Bacillus stearothermophilus]	89	80	1089
205	13	7875	7405	gi11216338	ORF for G15 ribosomal protein [Bacillus subtilis]	89	76	471
205	32	16152	15823	gi11165303	L3 [Bacillus subtilis]	89	80	330
270	3	2407	2207	gi1141902	arsenate reductase (EC 1.1.1.1) - Staphylococcus xylosus plasmid pSX267	89	81	201
395	2	157	672	gi11520574	glutamate racemase [Staphylococcus haemolyticus]	89	80	316
494	1	3	819	gi11396259	protease [Staphylococcus epidermidis]	89	77	837
510	1	1	444	gi1140046	phosphoglucose isomerase A (AA 1-449) [Bacillus stearothermophilus] trIS15936 NUB5A glucose-6-phosphate isomerase (EC 5.3.1.9) A - cillus stearothermophilus	89	74	444
615	1	2124	1210	gi11303812	YqeV [Bacillus subtilis]	89	74	915
841	1	18	341	gi11165303	L3 [Bacillus subtilis]	89	80	324
1111	1	352	813	gi1147146	thermonuclease [Staphylococcus intermedius]	89	70	462
1875	1	2	256	gi11205108	ATP-dependent protease binding subunit [Haemophilus influenzae]	89	82	255
2963	1	11	367	gi11467458	cell division protein [Bacillus subtilis]	89	83	357

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match	Match gene name	% sim	% ident	Length (nt)
3020	1	90	362	gi1239988	gi1239988	hypothetical protein [Bacillus subtilis]	89	66	273
3565	1	2	400	gi1256635	gi1256635	dihydroxy-acid dehydratase [Bacillus subtilis]	89	75	399
3586	1	105	314	gi1580832	gi1580832	ATP synthase subunit gamma [Bacillus subtilis]	89	82	210
3629	1	794	399	gi11009366	gi11009366	Respiratory nitrate reductase [Bacillus subtilis]	89	78	396
3688	1	2	400	gi1146206	gi1146206	glutamate dehydrogenase [Bacillus subtilis]	89	75	399
3699	1	794	399	gi1133950	gi1133950	large subunit of NADH-dependent glutamate synthase [Plectonoma boryanum]	89	75	396
4016	1	428	216	gi11009366	gi11009366	Respiratory nitrate reductase [Bacillus subtilis]	89	71	213
4177	1	471	301	gi1149426	gi1149426	putative [Lactococcus lactis]	89	76	171
4436	1	601	302	gi11022725	gi11022725	unknown [Staphylococcus haemolyticus]	89	80	300
4635	1	320	162	gi11022725	gi11022725	unknown [Staphylococcus haemolyticus]	89	73	159
2	2	1310	2676	gi1520754	gi1520754	putative [Bacillus subtilis]	88	76	1347
42	2	308	848	sr1423211.ATA	sr1423211.ATA	CATALASE (EC 1.11.1.6)	88	76	381
53	5	6389	4722	gi1474177	gi1474177	alpha-D-1,4-glucosidase [Staphylococcus xylosum]	88	80	1668
56	16	18018	18617	gi1467411	gi1467411	recombination protein [Bacillus subtilis]	88	77	600
60	3	376	843	gi1666116	gi1666116	glucose kinase [Staphylococcus xylosum]	88	77	468
70	2	1283	1245	gi144095	gi144095	replication initiator protein [Listeria monocytogenes]	88	74	339
82	8	11514	12719	gi1460663.A606	gi1460663.A606	translation elongation factor Tu - Bacillus subtilis	88	79	1206
103	7	4179	4391	gi1167181	gi1167181	serine/threonine kinase receptor [Brassica napus]	88	77	213
114	8	7712	8232	gi11022726	gi11022726	unknown [Staphylococcus haemolyticus]	88	72	501
118	2	308	2011	gi11303804	gi11303804	YgeQ [Bacillus subtilis]	88	77	1704
141	3	657	1136	gi11405446	gi11405446	transketolase [Bacillus subtilis]	88	72	480
148	7	5871	6116	gi1118002	gi1118002	dihydropterate synthase [Staphylococcus haemolyticus]	88	78	246
165	3	1428	2231	gi140053	gi140053	phenylalanyl-tRNA synthetase alpha subunit [Bacillus subtilis]	88	80	804
205	28	15027	14185	gi1165306	gi1165306	phenylalanyl-tRNA synthetase alpha subunit [Bacillus subtilis]	88	82	843
225	1	1569	898	gi11303840	gi11303840	YqfS [Bacillus subtilis]	88	78	672
235	1	2	1975	gi1452109	gi1452109	valyl-tRNA synthetase [Bacillus subtilis]	88	76	1974

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
339	3	2060	1566	gi1118002	dihydropterate synthase [Staphylococcus haemolyticus]	88	73	495
443	4	4325	2928	gi1558559	pyrimidine nucleoside phosphorylase [Bacillus subtilis]	88	73	1398
532	1	3	419	gi1143797	valyl-tRNA synthetase [Bacillus stearothermophilus] sp P1931 SVV_BACST VALYL-TRNA SYNTHETASE (EC 6.1.1.9) VALINE--TRNA LIGASE [VALAS]	88	78	417
534	3	2504	2968	gi153049	mammot-specific enzyme-III [Staphylococcus carnosus] pir JQ0088 JQ0088 phosphotransferase system enzyme II (EC 7.1.69) mannitol-specific factor III - Staphylococcus carnosus sp P17876 PTHA_STACA PTS SYSTEM, MANNITOL-SPECIFIC IIA COMPONENT EIIA-MTLI (	88	82	465
705	2	584	399	gi1710018	nitrate reductase (nirB) [Bacillus subtilis]	88	70	186
1000	2	1824	1309	gi11022726	unknown [Staphylococcus haemolyticus]	88	78	516
1299	1	587	324	gi1401786	phosphomannomutase [Mycoplasma pirum]	88	55	264
1341	2	170	400	gi139963	ribosomal protein L20 (AA 1-119) [Bacillus stearothermophilus] tr S05348 RS820 ribosomal protein L20 - Bacillus stearothermophilus	88	71	174
1386	1	41	214	gi1847154 8471	signal recognition particle 54K chain homolog Ffh - Bacillus subtilis	88	73	351
1386	2	183	533	gi11037154 8471	signal recognition particle 54K chain homolog Ffh - Bacillus subtilis	88	73	351
2949	1	704	399	gi1535350	CodX [Bacillus subtilis]	88	73	306
2984	1	5	169	gi1218277	10-acetylserine(thiol) lyase [Spinacia oleracea]	88	70	165
3035	1	1	178	gi1493083	dihydroxyacetone kinase [Citrobacter freundii]	88	67	138
3089	1	3	152	gi1606055	ORF_E746 [Escherichia coli]	88	88	150
3917	1	817	410	gi1143378	pyruvate decarboxylase (E-1) beta subunit [Bacillus subtilis]	88	77	408
4189	1	680	342	gi11405454	aconitase [Bacillus subtilis]	88	82	339
4201	1	734	369	gi1515938	glutamate synthase (ferredoxin) [Synecocystis sp.] pir S46957 S46957 glutamate synthase (ferredoxin) (EC 1.4.7.1) - ynechocystis sp.	88	84	366
4274	1	1	336	gi1515938	glutamate synthase (ferredoxin) [Synecocystis sp.] pir S46957 S46957 glutamate synthase (ferredoxin) (EC 1.4.7.1) - ynechocystis sp.	88	84	336
4308	1	784	399	gi11146206	glutamate dehydrogenase [Bacillus subtilis]	88	71	396
2	5	4570	6000	gi1535350	CodX [Bacillus subtilis]	87	70	1431
52	8	6781	6482	gi11064791	function unknown [Bacillus subtilis]	87	66	300

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
73	3	1584	2480	gi142992	glycerol kinase (glpk) (EC 2.7.1.30) (Bacillus subtilis) pir B45868 B45868 glycerol kinase (glpk) (EC 2.7.1.30) - Bacillus subtilis sp P18157 GLPK_BACSU GLYCEROL KINASE (EC 2.7.1.30) (ATP:GLYCEROL -PHOSPHOTRANSFERASE) (GLYCEROKINASE) (GK)	87	72	897
98	12	8813	9100	gi1467433	unknown (Bacillus subtilis)	87	62	288
124	4	4265	2988	gi1556886	serine hydroxymethyltransferase (Bacillus subtilis) pir S49363 S49363 serine hydroxymethyltransferase - Bacillus subtilis	87	77	1278
124	6	4457	4032	gi1556883	unknown (Bacillus subtilis)	87	66	426
148	5	3741	4559	gi1467460	unknown (Bacillus subtilis)	87	70	819
164	13	12710	13810	gi139954	if2 (aa 1-741) (Bacillus stearothermophilus)	87	72	1101
177	2	1164	2126	gi1467385	unknown (Bacillus subtilis)	87	78	1023
199	1	1982	1158	gi1443527	iron-sulfur protein (Bacillus subtilis)	87	77	825
199	2	4717	2933	pir A27763 A277	succinate dehydrogenase (EC 1.3.99.1) flavoprotein - Bacillus subtilis	87	80	1785
205	23	11782	11543	gi11044972	ribosomal protein l29 (Bacillus subtilis)	87	75	669
205	25	11325	12607	gi11165309	l53 (Bacillus subtilis)	87	70	927
222	1	2033	1107	gi11177249	rec23 gene product (Bacillus subtilis)	87	80	303
236	3	1635	1333	gi11146198	ferredoxin (Bacillus subtilis)	87	77	294
246	5	2585	2292	gi1467373	ribosomal protein S18 (Bacillus subtilis)	87	72	768
260	2	4189	3422	gi11161382	lceC (Staphylococcus epidermidis)	87	80	696
320	3	1696	2391	gi1312443	carbamoyl-phosphate synthase (glutamine-hydrolyzing) (Bacillus aldoityticus)	87	80	696
380	4	1165	1383	gi1142570	ATP synthase c subunit (Bacillus firmus)	87	80	219
414	4	200	1073	gi1467386	thiophen and furan oxidation (Bacillus subtilis)	87	77	174
425	12	1003	794	gi11046166	pilin repressor (Mycoplasma genitalium)	87	69	210
448	1	1255	722	gi1405134	acetate kinase (Bacillus subtilis)	87	75	534
480	1	1	711	gi1142659	ATP synthase alpha subunit (Bacillus megaterium)	87	78	711
481	1	2	352	sp Q06797 KUL_B	150S RIBOSOMAL PROTEIN L1 (BL1)	87	72	351
677	2	359	955	gi1460911	fructose-bisphosphate aldolase (Bacillus subtilis)	87	78	597
677	3	934	1284	gi1460911	fructose-bisphosphate aldolase (Bacillus subtilis)	87	78	351
876	1	3	452	gi1146247	asparaginyl-tRNA synthetase (Bacillus subtilis)	87	79	450

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
1376	1	426	214	gi1065555	[F46H6.4 gene product (Caenorhabditis elegans)]	87	75	213
2206	1	3	374	gi1215098	[excisionase (Bacteriophage 154a)]	87	72	372
2918	1	3	290	gi1508979	[GTP-binding protein (Bacillus subtilis)]	87	69	288
3081	2	126	308	gi1467399	[IMP dehydrogenase (Bacillus subtilis)]	87	72	183
3535	1	3	401	gi11405454	[aconitase (Bacillus subtilis)]	87	80	399
4218	1	547	275	gi1603769	[HsdR protein, urocanase (Bacillus subtilis)]	87	73	273
4	8	10427	8736	gi1603769	[HsdR protein, urocanase (Bacillus subtilis)]	86	72	1692
22	6	4190	3738	gi1410515	[urease beta subunit (Staphylococcus xylosum)]	86	73	453
54	2	2480	1572	gi1289287	[UDP-glucose pyrophosphorylase (Bacillus subtilis)]	86	70	909
124	3	2336	1713	gi1556687	[uracil phosphoribosyltransferase (Bacillus subtilis)]	86	74	624
148	3	1349	3448	gi1467458	[uracil phosphoribosyltransferase - Bacillus subtilis]	86	75	2100
148	4	3638	3859	gi1467460	[cell division protein (Bacillus subtilis)]	86	73	222
152	3	1140	2086	gi1377835	[unknown (Bacillus subtilis)]	86	75	747
164	18	17347	19467	gi1184680	[pyruvate decarboxylase E-1 alpha subunit (Bacillus subtilis)]	86	72	2121
180	2	554	1159	gi1143467	[polynucleotide phosphorylase (Bacillus subtilis)]	86	80	606
205	3	2966	2592	gi1142464	[ribosomal protein S4 (Bacillus subtilis)]	86	77	175
205	26	13364	11290	gi140107	[ribosomal protein L17 (Bacillus subtilis)]	86	75	375
246	7	3463	3140	gi1467375	[ribosomal protein L22 (Bacillus stearothermophilus)]	86	70	324
299	3	1196	1540	gi139656	[ribosomal protein L22 - Bacillus stearothermophilus]	86	70	345
299	7	3884	4345	gi1467440	[rpoVG gene product (Bacillus megaterium)]	86	78	462
304	5	2170	2523	gi1666983	[phosphoribosylpyrophosphate synthetase (Bacillus subtilis)]	86	65	354
310	2	1487	1676	gi1177684	[synthetase (AA 1-37) (Bacillus subtilis)]	86	71	192
317	5	2086	3405	gi1487434	[putative ATP binding subunit (Bacillus subtilis)]	86	78	1320
319	2	1489	1109	gi1118003	[chorismate mutase (Staphylococcus xylosum)]	86	77	381
358	2	2124	3440	gi1146219	[isocitrate dehydrogenase (Bacillus subtilis)]	86	73	1317
					[dihydropyrimidin aldolase (Staphylococcus haemolyticus)]			
					[28.2% of identity to the Escherichia coli GTP-binding protein Era; putative (Bacillus subtilis)]			

TABLE 2



S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% iden	length (nt)
404	2	1015	2058	gi 1303817	YqfA (Bacillus subtilis)	86	78	1044
581	2	661	452	gi 400556	phoP gene product (Bacillus subtilis)	86	71	210
642	2	338	1075	gi 1176399	epiF (Staphylococcus epidermidis)	86	72	738
770	1	622	347	gi 143328	phoP protein (put.): putative (Bacillus subtilis)	86	69	276
865	1	1777	890	gi 1146247	asparaginyl-tRNA synthetase (Bacillus subtilis)	85	74	848
868	2	963	1133	gi 1002911	transmembrane protein (Saccharomyces cerevisiae)	85	69	171
904	1	1	162	gi 1303912	YqjW (Bacillus subtilis)	86	72	162
989	1	35	433	gi 1303993	YqjL (Bacillus subtilis)	86	76	399
1212	1	296	150	gi 414014	ipa-900 gene product (Bacillus subtilis)	86	70	147
1323	1	2	148	gi 40341	pyruvate dehydrogenase (lipoamide) (Bacillus stearothermophilus) ir S10798 DEBSPF pyruvate dehydrogenase (lipoamide) (EC 1.2.4.1) pha chain - Bacillus stearothermophilus	86	75	147
3085	2	540	310	gi 1354211	PEP112-like protein (Bacillus subtilis)	86	86	231
1847	1	1	228	gi 296464	ATPase (Lactococcus lactis)	86	63	278
4487	1	476	240	gi 1022726	unknown (Staphylococcus haemolyticus)	86	73	237
4583	1	372	187	gi 1022725	unknown (Staphylococcus haemolyticus)	86	79	186
25	5	4287	5019	gi 1502421	3-ketoacyl-acyl carrier protein reductase (Bacillus subtilis)	85	64	753
56	21	10627	29395	gi 1408507	pyrimidine nucleoside transport protein (Bacillus subtilis)	85	69	1213
68	2	332	1192	gi 467376	unknown (Bacillus subtilis)	85	74	861
73	2	880	1707	gi 142992	glycerol kinase (glpK) (EC 2.7.1.30) (Bacillus subtilis) pir B45868 B45868 glycerol kinase (EC 2.7.1.30) - Bacillus subtilis sp P18157 GLPK_BACSU GLYCEROL KINASE (EC 2.7.1.30) (ATP:GLYCEROL -PHOSPHOTRANSFERASE) (GLYCEROKINASE) (GK)	85	72	828
106	4	1505	3490	gi 143766	(ttrSV) (EC 6.1.1.3) (Bacillus subtilis)	85	74	1986
128	2	1153	2202	gi 111924	glyceraldhyde-3-phosphate dehydrogenase (Clostridium pasteurianum) pir S14254 S14254 glyceraldehyde-3-phosphate dehydrogenase (EC 2.1.12) - Clostridium pasteurianum	85	75	1050
129	4	6466	5252	gi 1064807	ORTHINE AMINOTRANSFERASE (Bacillus subtilis)	85	73	1215
138	6	3475	5673	gi 1072419	glcB gene product (Staphylococcus carnosus)	85	74	2199
189	1	2	169	gi 467385	unknown (Bacillus subtilis)	85	65	168

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
205	15	8624	8106	gi11044981	ribosomal protein S5 [Bacillus subtilis]	85	75	519
205	20	10928	10596	pirA02819/RSB5	ribosomal protein L24 - Bacillus stearothermophilus	85	72	333
220	6	6490	6101	gi148980	secA gene product [Bacillus subtilis]	85	66	390
231	4	4877	3159	gi11002520	MutS [Bacillus subtilis]	85	70	1719
243	9	8013	8783	gi1414011	lpa-87 gene product [Bacillus subtilis]	85	72	771
249	2	5894	3186	gi11405454	aconitase [Bacillus subtilis]	85	73	2709
302	1	140	475	gi140173	homolog of E. coli ribosomal protein L21 [Bacillus subtilis] ir S18439 S18439 Ribosomal protein L21 - Bacillus subtilis p P26908 RL21_BACSU 50S RIBOSOMAL PROTEIN L21 (BL20)	85	72	336
333	1	5445	2968	gi1442360	ClpC adenosine triphosphatase [Bacillus subtilis]	85	69	2478
364	6	6062	8196	gi1871784	Clp-like ATP-dependent protease binding subunit [Bos taurus]	85	68	2115
448	2	1992	1339	gi1405134	acetate kinase [Bacillus subtilis]	85	68	654
747	1	1251	853	gi11373157	orf-X; hypothetical protein. Method: conceptual translation supplied by author [Bacillus subtilis]	85	71	399
886	2	159	467	gi1541768	hemolysin [Versinia enterocolitica]	85	55	309
1089	1	1208	606	pir 847154 8471	signal recognition particle 54K chain homolog Ffh - Bacillus subtilis	85	71	603
1163	1	816	409	gi1301155	diaminopimelate decarboxylase [Bacillus methanolicus] sp P41023 DCDA_BACMT DIAMINOPIMELATE DECARBOXYLASE (EC 4.1.1.20) DAP DECARBOXYLASE	85	62	408
1924	1	487	251	gi1215098	excisionase [Bacteriophage 154a]	85	73	237
2932	1	716	390	gi11041099	Pyruvate Kinase [Bacillus licheniformis]	85	71	387
3030	1	3	275	gi142370	pyruvate formate-lyase (AA 1-760) [Escherichia coli] ir S01788 S01788 formate C-acetyltransferase (EC 2.3.1.54) - Chericchia coli	85	74	273
3111	1	595	299	gi163568	limb deformity protein [Gallus gallus]	85	85	297
3778	1	630	316	gi1391840	beta-subunit of HDT [Pseudomonas fragilis]	85	67	315
3815	1	1	387	gi11204472	type I restriction enzyme ECOR124/3 I M protein [Haemophilus influenzae]	85	56	387
4042	1	3	386	gi110178	formate acetyltransferase [Chlamydomonas reinhardtii] ir S24997 S24997 formate C-acetyltransferase (EC 2.3.1.54) - Chlamydomonas reinhardtii	85	70	384
4053	1	35	340	gi11204472	type I restriction enzyme ECOR124/3 I M protein [Haemophilus influenzae]	85	56	306
4108	1	2	181	gi11072418	glcA gene product [Staphylococcus carnosus]	85	61	180
4300	1	575	330	gi1151932	fructose enzyme II [Rhodospirillum rubrum]	85	59	246

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4392	1	627	355	gi11022725	unknown [Staphylococcus haemolyticus]	85	74	273
4408	1	2	235	gi11471744	Cip-like ATP-dependent protease binding subunit [Bos taurus]	85	62	234
4430	1	578	291	gi11009366	Respiratory nitrate reductase [Bacillus subtilis]	85	68	288
4555	1	2	253	gi1450688	hadM gene of Ecopzi gene product [Escherichia coli] pir153437 [S38437 hadM protein - Escherichia coli pir1509629 [S09629 hypothetical protein A - Escherichia coli] (S08 40-520)]	85	52	252
4611	1	481	242	gi11256635	glyoxy-acid dehydratase [Bacillus subtilis]	85	65	240
4	10	11061	10591	gi146982	foaB gene product [Staphylococcus epidermidis]	84	68	511
13	2	1348	1172	gi1142450	leuC protein [Bacillus subtilis]	84	56	177
16	4	1803	4652	gi11277198	DNA repair protein [Deinococcus radiodurans]	84	67	2850
22	3	1535	1128	gi1511069	Orf1 [Staphylococcus xylosum]	84	73	408
23	7	5055	5306	gi1603320	Yer082p [Saccharomyces cerevisiae]	84	61	254
53	11	11597	11145	gi11303948	Yqjw [Bacillus subtilis]	84	68	453
53	12	114059	11270	gi1142613	branched chain alpha-keto acid dehydrogenase E2 [Bacillus subtilis]	84	71	1270
70	1	1332	982	gi146647	ORF (repB) [Staphylococcus aureus]	84	68	351
73	4	2512	4311	gi1142593	glycerol-3-phosphate dehydrogenase (gldP) (EC 1.1.99.5) [Bacillus subtilis]	84	74	1800
94	7	4324	6096	gi1467427	methylonyl-tRNA synthetase [Bacillus subtilis]	84	66	1773
100	9	9501	8680	gi11340128	ORF1 [Staphylococcus aureus]	84	78	822
117	3	1934	3208	gi11237019	Srb [Bacillus subtilis]	84	68	1275
148	6	4720	5670	gi1467462	cysteine synthetase A [Bacillus subtilis]	84	69	951
152	4	2064	2456	gi1141377	pyruvate decarboxylase (E-1) alpha subunit [Bacillus subtilis]	84	70	393
				pir1816718 [H8SPA pyruvate dehydrogenase (lipoamide) (EC 1.2.4.11) lpha chain - Bacillus subtilis]				
169	7	3634	3861	gi11001342	hypothetical protein [Synecocystis sp.1]	84	66	228
171	4	2992	2657	gi1517475	D-amino acid transaminase [Staphylococcus haemolyticus]	84	71	330
186	6	6941	6216	gi1467475	unknown [Bacillus subtilis]	84	70	726
205	9	6261	5692	gi1216340	ORF for adenylate kinase [Bacillus subtilis]	84	71	570
224	2	915	1391	gi1288269	beta-fructofuranosidase [Staphylococcus xylosum]	84	70	477

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
251	1	92	388	gi 1303790	ygeI [Bacillus subtilis]	84	65	297
282	3	1526	2836	gi 143040	glutamate-1-semialdehyde 2,1-aminotransferase [Bacillus subtilis] pir D42728 D42728 glutamate-1-semialdehyde 2,1-aminotransferase (EC 4.3.8) - Bacillus subtilis	84	75	1311
307	5	3118	2959	gi 1070014	protein-dependent [Bacillus subtilis]	84	62	180
320	4	2343	4229	gi 143390	[carbamy] phosphate synthetase [Bacillus subtilis]	84	70	1987
372	1	3	286	gi 1022725	unknown [Staphylococcus haemolyticus]	84	70	294
413	2	2201	1341	gi 1256146	ybbQ [Bacillus subtilis]	84	65	861
439	1	3	392	gi 1046173	osmotically inducible protein [Mycoplasma genitalium]	84	53	390
461	3	1362	2270	gi 40211	threonine synthase (thiC) [AA 1-352] [Bacillus subtilis] ir A25364 A25364 pir D42728 D42728 glutamate-1-semialdehyde 2,1-aminotransferase (EC 4.2.99.2) - Bacillus subtilis	84	69	509
487	1	3	299	gi 1144531	integrin-like protein alpha Intp [Candida albicans]	84	48	297
491	2	624	905	pir 508564 8385	ribosomal protein S9 - Bacillus stearothermophilus	84	69	282
491	3	836	1013	pir 508564 8385	ribosomal protein S9 - Bacillus stearothermophilus	84	77	198
548	1	3	341	gi 111271	[uracil] permease [Bacillus caldolyticus]	84	74	339
728	2	2701	1748	gi 912445	DNA polymerase [Bacillus caldotenax]	84	68	954
769	1	3	257	gi 1510953	isobutylamine biosynthesis protein N [Methanococcus jannaschii]	84	38	255
974	1	308	156	gi 1405454	lactonase [Bacillus subtilis]	84	57	151
957	1	3	395	gi 143402	recombination protein (rtg start codon) [Bacillus subtilis] gi 1303923 RecN [Bacillus subtilis]	84	68	393
975	1	3	452	gi 885914	ClpB [Synecoccus sp.]	84	70	450
1585	1	3	257	gi 1510140	liqendopeptidase F [Lactococcus lactis]	84	56	255
2954	1	3	323	gi 603769	HutU protein, uracinate [Bacillus subtilis]	84	73	321
2996	1	650	348	gi 18178	formate acetyltransferase [Chlamydomonas reinhardtii] ir S24957 S24957 formate C-acetyltransferase (EC 2.3.1.34) - Chlamydomonas reinhardtii	84	65	303
3766	1	737	375	gi 1517205	167 kDa Myosin-crossreactive streptococcal antigen [Streptococcus yogenes]	84	72	363
4022	1	2	169	gi 1146206	glutamate dehydrogenase [Bacillus subtilis]	84	54	168
4058	1	620	312	gi 151912	fructose enzyme II [Rhodospirillum rubrum]	84	71	309
4108	2	106	351	gi 1072418	glcA gene product [Staphylococcus carnosus]	84	77	246

TABLE 2

S. aureus - putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
4183	1	3	308	gi1603769	[hutU protein, urocanase (Bacillus subtilis)]	84	72	306
4726	1	55	234	gi146208	[glutamate synthase large subunit (EC 2.6.1.53) (Escherichia coli) pifA29617/A29617 glutamate synthase (NADPH) (EC 1.4.1.13) large chain - Escherichia coli]	84	73	183
22	4	2043	1576	gi1393297	[urease accessory protein (Bacillus sp.)]	83	64	468
53	13	134722	13745	gi1142612	[branched chain alpha-keto acid dehydrogenase E1-beta (Bacillus subtilis)]	83	68	978
57	16	13357	12872	gi1143132	[lactate dehydrogenase (AC 1.1.1.27) (Bacillus caldolyticus) pifB29704/B29704 L-lactate dehydrogenase (EC 1.1.1.27) - Bacillus caldolyticus]	83	66	486
66	3	3119	2274	gi1103894	[YqkM (Bacillus subtilis)]	83	63	846
66	5	6118	4643	gi11212730	[YqkM (Bacillus subtilis)]	83	68	1476
70	3	1864	1523	gi144095	[replication initiator protein (Listeria monocytogenes)]	83	73	342
90	1	377	1429	gi1155571	[alcohol dehydrogenase I (adhA) (EC 1.1.1.1) (Zymomonas mobilis) pifA35260/A35260 alcohol dehydrogenase (EC 1.1.1.1) I - Zymomonas mobilis]	83	70	1053
95	2	708	2162	gi1506381	[phospho-beta-glucosylase (Bacillus subtilis)]	83	70	1455
137	1	68	694	gi1467391	[initiation protein of replication (Bacillus subtilis)]	83	77	627
140	4	3209	2742	gi1634107	[kdpB (Escherichia coli)]	83	65	468
142	3	3468	2989	gi11212776	[lumazine synthase (b-subunit) (Bacillus anthracis)]	83	69	480
161	12	5749	6696	gi1903307	[ORF75 (Bacillus subtilis)]	83	64	948
164	9	9880	11070	gi149316	[ORF2 gene product (Bacillus subtilis)]	83	66	1191
164	14	16168	14546	gi1580902	[ORF6 gene product (Bacillus subtilis)]	83	60	399
170	2	3144	2467	gi1520844	[orf4 (Bacillus subtilis)]	83	64	674
186	2	2029	1370	gi1289284	[cysteine-tRNA synthetase (Bacillus subtilis)]	83	72	650
205	14	7822	7607	gi1216337	[ORF for L30 ribosomal protein (Bacillus subtilis)]	83	74	216
237	6	3693	4540	gi11510488	[imidazoleglycerol-phosphate synthase (cyclase) (Methanococcus jannaschii)]	83	60	858
301	1	985	638	gi1467419	[unknown (Bacillus subtilis)]	83	65	348
302	4	1421	2743	gi1508979	[GTP-binding protein (Bacillus subtilis)]	83	68	1323
321	4	3933	3571	gi139844	[fumarate (citG) (aa 1-462) (Bacillus subtilis)]	83	68	363
367	1	2	352	gi11039479	[ORF1 (Lactococcus lactis)]	83	54	351

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
387	1	3	662	gi 806281	URA polymerase I [Bacillus stearothermophilus]	83	70	660
527	2	916	1566	gi 396259	protease [Staphylococcus epidermidis]	83	67	651
533	1	355	179	g 142455	alanine dehydrogenase (EC 1.4.1.1) [Bacillus stearothermophilus]	83	66	177
					alanine dehydrogenase (EC 1.4.1.1) - Bacillus stearothermophilus			
536	4	1617	1438	g 143366	adenylosuccinate lyase (PDB-B) [Bacillus subtilis]	83	67	180
					adenylosuccinate lyase (EC 4.3.2.2) - Bacillus subtilis			
652	1	2	859	gi 520753	DNA topoisomerase I [Bacillus subtilis]	83	72	858
774	2	200	361	gi 1522665	M. jannaschii predicted coding region MJCT28 [Methanococcus jannaschii]	83	58	162
897	1	120	226	gi 1064807	ORTHONINE AMINOTRANSFERASE [Bacillus subtilis]	83	76	177
1213	1	3	491	gi 289288	lexA [Bacillus subtilis]	83	67	489
2529	1	296	150	gi 143786	cryptophanyl-tRNA synthetase (EC 6.1.1.2) [Bacillus subtilis]	83	69	147
					pir JF048: YMB5 cryptophan-tRNA ligase (EC 6.1.1.2) - Bacillus ubtilla			
2873	1	649	326	gi 1109687	ProZ [Bacillus subtilis]	83	58	324
1009	1	718	166	gi 1492532	UMP_0294 [Escherichia coli]	83	65	363
3035	2	45	305	gi 950062	hypothetical yeast protein 1 [Mycoplasma capricolum] pir 548578 548578	83	59	281
					hypothetical protein - Mycoplasma capricolum SCC3 (fragment)			
3406	1	67	307	gi 1353197	chloroquin reductase [Eubacterium acidaminophilum]	83	61	241
4454	1	510	271	gi 397526	clumping factor [Staphylococcus aureus]	83	78	270
4570	1	444	223	gi 1322726	unknown [Staphylococcus haemolyticus]	83	74	222
4654	1	97	261	gi 1072419	glcB gene product [Staphylococcus carnosus]	83	79	165
16	2	295	1191	gi 153854	juv402 protein [Streptococcus pneumoniae]	82	67	897
16	3	1393	1798	gi 153854	juv402 protein [Streptococcus pneumoniae]	82	70	606
38	12	9644	8724	gi 1204400	N-acetylneuraminic lyase [Haemophilus influenzae]	82	58	921
42	4	988	2019	gi 861192	catalase [Bacteroides fragilis]	82	70	1032
51	6	2550	3489	gi 143607	sporulation protein [Bacillus subtilis]	82	69	900
56	11	12270	13925	gi 39431	oligo-1,6-glucosidase [Bacillus cereus]	82	60	1656
56	15	17673	18014	gi 467410	unknown [Bacillus subtilis]	82	66	342
61	2	881	3313	gi 143148	transfer RNA-Leu synthetase [Bacillus subtilis]	82	70	2433

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
82	7	9162	11318	gi148240	elongation factor G (AA 1-691) [Thermus aquaticus thermophilus] tr[S1592]EPTWG translation elongation factor G - Thermus aquaticus p[P1353]EFC_THETH_ELONGATION_FACTOR_G [EF-G]	82	64	2157
85	2	5470	3260	gi1143369	phosphoribosylformyl glycine synthetase II (pur-Q) [Bacillus subtilis]	82	66	2211
102	6	3662	5180	gi11256635	dihydroxy-acid dehydratase [Bacillus subtilis]	82	65	1719
117	4	3242	3493	gi11447541A471	orf15' of Ffn - Bacillus subtilis	82	53	252
128	6	4377	5933	gi1460258	phosphoglycerate mutase [Bacillus subtilis]	82	66	1557
129	2	1229	2182	gi1403373	glycerophosphoryl diester phosphodiesterase [Bacillus subtilis] pic[S1725]S17251 glycerophosphoryl diester phosphodiesterase - acillus subtilis	82	62	953
170	1	2	1441	gi1137831	unknown [Bacillus subtilis]	82	67	1440
177	1	3	1094	gi1467386	thiophen and furan oxidation [Bacillus subtilis]	82	65	1092
184	4	3572	4039	gi1153566	ORF (19K protein) [Enterococcus faecalis]	82	59	468
189	8	4455	4225	gi11001878	CapL protein [Listeria monocytogenes]	82	73	231
206	119	21166	20707	gi1473916	lipopeptide antibiotic iturin A [Bacillus subtilis] ap[P93144]IP14_BACSU LIPPEPTIDE ANTIBIOTICS ITURIN A AND SURFACTIN BIOSYNTHESIS PROTEIN	82	50	660
221	2	805	1722	gi1517205	67 kDa Myosin-crossreactive streptococcal antigen [Streptococcus yoenes]	82	63	918
223	4	3866	3651	gi1439619	[Salmonella typhimurium] IS200 insertion sequence from SAR17, artial 1, gene product [Salmonella typhimurium]	82	69	216
260	3	5207	4296	gi11161381	IcaB [Staphylococcus epidermidis]	82	61	912
315	3	4864	2855	gi1143397	guinol oxidase [Bacillus subtilis]	82	67	2010
321	110	8520	7945	gi1143981	ORF5: This ORF includes a region (ae23-103) containing a potential non- sulphur centre homologous to a region of Rhodospirillum rubrum nd Chromatium vinosum; putative [Bacillus stearothermophilus] pic[PQ0295]PQ0299 hypothetical protein 5 (glda 3' region) -	82	62	576
331	3	1055	1342	gi1436574	ribosomal protein L1 [Bacillus subtilis]	82	71	288
370	2	262	618	gi11303793	YqeL [Bacillus subtilis]	82	59	357
404	4	3053	4024	gi11303821	YqfE [Bacillus subtilis]	82	68	972
405	4	4440	3073	gi11303913	YqhX [Bacillus subtilis]	82	67	1368
436	3	4096	2864	gi1149521	tryptophan synthase beta subunit [Bacillus subtilis] pic[S15129]S15129 tryptophan synthase (EC 4.2.1.20) beta chain - actinococcus lactis subsp. lactis	82	67	1233

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
441	4	3394	2573	gi142952	glycerate dehydrogenase [Bacillus stearothermophilus]	82	67	822
444	12	10415	11227	gi1264354	spore germination and vegetative growth protein [Haemophilus influenzae]	82	67	813
446	1	3	191	gi143387	aspartate transcarbamylase [Bacillus subtilis]	82	66	189
462	3	1007	1210	gi142521	deoxyribodipyridine photolase [Bacillus subtilis] p147192[AJ7192] uvrB protein - Bacillus subtilis sp[P14951][UVRC_BACSU EXONUCLEASE ABC SUBUNIT C]	82	64	204
537	1	1560	784	gi185767	UDP-N-acetylglucosamine 1-carboxyvinyltransferase [Bacillus subtilis]	82	61	777
680	2	407	700	gi1426472	secE gene product [Staphylococcus carnosus]	82	69	294
724	2	565	386	gi143373	phosphoribosyl aminoimidazole carboxy formyl ormyltransferase/inosine monophosphate cyclohydrolase (PUR-H13) [Bacillus subtilis]	82	68	180
763	1	422	213	gi1457458	cell division protein [Bacillus subtilis]	82	35	210
818	1	564	281	gi1064787	function unknown [Bacillus subtilis]	82	69	282
858	1	175	1176	gi143043	uroporphyrinogen decarboxylase [Bacillus subtilis] p147045[B47045] uroporphyrinogen decarboxylase [EC 4.1.1.37] - actillus subtilis	82	71	1002
895	1	3	599	gi11027507	ATP binding protein [Borrelia burgdorferi]	82	72	597
939	1	10	399	gi1143795	transfer RNA-Tyr synthetase [Bacillus subtilis]	82	60	390
961	1	1	306	gi1577647	gamma-hemolysin [Staphylococcus aureus]	82	69	306
1392	1	307	155	gi1146974	NHJ-dependent NAD synthetase [Escherichia coli]	82	71	153
1317	1	49	375	gi1407908	EIIICr [Staphylococcus xylosum]	82	72	327
1341	1	1	150	gi139962	ribosomal protein L35 (AA 1-66) [Bacillus stearothermophilus] lr[S05347][S5835 ribosomal protein L35 - Bacillus stearothermophilus]	82	68	150
2990	2	567	349	gi1534855	ATPase subunit epsilon [Bacillus stearothermophilus] sp[P42009][ATPE_BACST ATP SYNTHASE EPSILON CHAIN [EC 3.6.1.34]	82	47	219
3024	1	45	226	gi1467402	unknown [Bacillus subtilis]	82	64	180
3045	1	276	139	gi1467335	ribosomal protein L9 [Bacillus subtilis]	82	60	138
3045	2	558	400	gi1467335	ribosomal protein L9 [Bacillus subtilis]	82	82	159
3091	1	474	238	gi1499335	secA protein [Staphylococcus carnosus]	82	78	237
3107	1	416	210	gi1546918	orfY 3' of comK [Bacillus subtilis, E26, Peptide Partial, 140 aa] p147046[2][S43612] hypothetical protein Y - Bacillus subtilis sp[P40398][YND_BACSU HYPOTHETICAL PROTEIN IN COMK 3'-REGION (ORF) FRAGMENT]	82	64	207

TABLE 2



S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
432	1	2	319	gi 42086	nitrate reductase alpha subunit [Escherichia coli] p P09152 NARG_ECOLI	82	75	318
					RESPIRATORY NITRATE REDUCTASE 1 ALPHA CHAIN (EC 7.99.4). (SUB 2-1247)			
23	3	3275	2574	gi 1199573	spSB [Sphingomonas sp.]	81	64	702
42	1	638	321	gi 466378	lysine specific permease [Escherichia coli]	81	59	318
48	5	4051	4350	gi 1045937	M. genitalium predicted coding region NG246 [Mycoplasma genitalium]	81	62	300
51	4	1578	2579	pir S16649 S166	dcIAC protein - Bacillus subtilis	81	55	1002
53	2	354	1494	gi 1303961	YQJ2 [Bacillus subtilis]	81	67	1131
53	8	9419	7971	gi 146930	6-phosphogluconate dehydrogenase [Escherichia coli]	81	66	1449
54	9	10757	13119	gi 143016	permease [Bacillus subtilis]	81	65	639
54	10	13360	11786	gi 143015	gluconate kinase [Bacillus subtilis]	81	64	1575
57	17	13983	13366	pir N25805 N258	L-lactate dehydrogenase (EC 1.1.1.27) - Bacillus subtilis	81	74	618
81	2	2708	2217	gi 1222302	N160 related protein [Haemophilus influenzae]	81	54	492
86	1	745	374	gi 414017	jha-93D gene product [Bacillus subtilis]	81	70	372
103	6	6438	4861	gi 971342	nitrate reductase beta subunit [Bacillus subtilis] sp P42176 NANH_BACSU	81	64	1578
					NITRATE REDUCTASE BETA CHAIN (EC 1.7.99.4)			
120	15	10845	12338	gi 1524392	GbsA [Bacillus subtilis]	81	67	1494
128	5	1676	4413	gi 143319	triiose phosphate isomerase [Bacillus megaterium]	81	64	719
131	9	10308	9280	gi 299163	alanine dehydrogenase [Bacillus subtilis]	81	68	1029
143	6	6088	5471	gi 439619	[Salmonella typhimurium] 15200 insertion sequence from SARA17, article 1, gene product [Salmonella typhimurium]	81	61	618
169	1	43	825	gi 697795	30S ribosomal protein [Pedococcus acidilactici] sp P49668 RS2_PEDAC 30S	81	65	783
					RIBOSOMAL PROTEIN S2			
230	1	450	226	gi 1125826	short region of weak similarity to tyrosine-protein kinase receptors in a fibronectin type III-like domain [Caenorhabditis elegans]	81	54	225
233	5	2000	2677	gi 467404	unknown [Bacillus subtilis]	81	63	678
241	2	3081	2149	gi 16510	succinate-CoA lyase (GDP-forming) [Arabidopsis thaliana] tr S30579 S30579	81	69	933
					succinate-CoA lyase (GDP-forming) (EC 6.2.1.4) pna chain - Arabidopsis thaliana (fragment)			
256	1	1	981	nit S09411 S094	spoIIIE protein - Bacillus subtilis	81	65	981
259	3	3752	7691	sp P28167 PF2_B	PROBABLE PEPTIDE CHAIN RELEASE FACTOR 2 (RP-2) (FRAGMENT)	81	65	1062

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
275	2	1728	3581	gi 726480	[L-glutamine-D-fructose-6-phosphate amidotransferase (Bacillus subtilis)]	81	68	1854
285	1	1466	735	gi 1204844	[H. influenzae predicted coding region HD594 (Haemophilus influenzae)]	81	63	732
296	1	99	1406	gi 467328	[adenylosuccinate synthetase (Bacillus subtilis)]	81	67	1308
302	9	5590	5689	gi 147485	[queA (Escherichia coli)]	81	64	100
317	2	1137	1376	gi 1354961	[resolvase (Transposon Tn917)]	81	57	240
343	2	1034	1342	gi 405955	[yeeD (Escherichia coli)]	81	60	109
360	2	1404	2471	gi 1204570	[aspartyl-tRNA synthetase (Haemophilus influenzae)]	81	67	1068
364	5	6351	5706	gi 1204652	[methylated-DNA-protein-cysteine methyltransferase (Haemophilus influenzae)]	81	63	546
372	2	1707	1135	gi 467416	[unknown (Bacillus subtilis)]	81	65	573
392	1	43	603	pir S09411 S094	[spoIIIE protein - Bacillus subtilis]	81	65	561
404	9	5252	6154	gi 606745	[Bex (Bacillus subtilis)]	81	65	903
426	2	1727	1119	gi 39453	[Manganese superoxide dismutase (Bacillus caldotenax) (ir S22053 S22053) superoxide dismutase (EC 1.15.1.1) (Mn) - Bacillus lichenex]	81	66	609
440	7	5053	5889	gi 137043 C170	[hypothetical protein II (loopII 3' region) - Salmonella typhimurium (fragment)]	81	57	237
625	3	1105	2070	gi 1262360	[protein kinase PknB (Mycobacterium leprae)]	81	56	966
754	2	504	1064	gi 1303902	[yqnW (Bacillus subtilis)]	81	71	561
842	1	86	430	gi 1405446	[transketolase (Bacillus subtilis)]	81	64	345
953	1	798	400	gi 1205429	[dipeptide transport ATP-binding protein (Haemophilus influenzae)]	81	57	399
961	2	252	401	gi 487686	[synergohymenotropic toxin (Staphylococcus intermedius) pir S46946 S46944 synergohymenotropic toxin - Staphylococcus intermedius]	81	72	150
1005	1	1	189	gi 1046138	[M. genitalium predicted coding region MG423 (Mycoplasma genitalium)]	81	43	189
1280	1	670	449	gi 559164	[helicase (Autographa californica nuclear polyhedrosis virus) ap P24307 V143.NPVAC.HELICASE]	81	43	222
3371	1	68	241	gi 1322245	[mavalonate pyrophosphate decarboxylase (Rattus norvegicus)]	81	62	174
3715	1	475	239	gi 537137	[ORF_188 (Escherichia coli)]	81	58	217
3904	1	2	325	gi 439619	[Salmonella typhimurium IS200 insertion sequence from SAR417, attC1-1, gene product (Salmonella typhimurium)]	81	68	324
3940	1	3	401	gi 1296464	[ATPase (Lactococcus lactis)]	81	69	399

TABLE 2

3. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3954	1	1	318	gi1224069	amidase [Morexella catarrhalis]	81	68	318
4049	1	337	170	gi1603768	MutI protein, imidazole-5-propionate hydrolase [Bacillus subtilis]	81	68	168
4209	1	1	324	gi1603768	glycerophosphoryl diester phosphodiesterase [Bacillus subtilis]	81	58	324
4371	1	627	322	gi1216677	indolepyruvate decarboxylase [Enterobacter cloacae] pif16033[S1603]	81	72	306
4387	1	19	228	gi1460689	indolepyruvate decarboxylase (EC 4.1.1.1) - nterobacter cloacae	81	59	210
4391	1	581	306	gi11524193	TVC [Thermactinomyces vulgaris]	81	67	276
4425	1	3	341	gi1143015	unknown [Mycobacterium tuberculosis]	81	66	339
9	1	593	847	gi11064786	glucuronate kinase [Bacillus subtilis]	80	62	747
17	1	544	311	gi1559164	function unknown [Bacillus subtilis]	80	40	234
45	2	1159	2448	gi11109684	helicase [Autographa californica nuclear polyhedrosis virus]	80	63	1290
45	5	4032	4733	gi11109687	ap194307[V143_NPVAC HELICASE]	80	55	702
54	8	10266	9502	gi1563952	ProV [Bacillus subtilis]	80	82	765
62	12	8852	7545	gi1854655	glucuronate permease [Bacillus licheniformis]	80	62	1308
62	14	8087	8683	gi1559713	Na/H antiporter system [Bacillus alcalophilus]	80	68	597
67	16	13781	14122	gi1305002	ORF 1100 [Homo sapiens]	80	65	342
70	13	11495	10296	gi11303995	ORF 4356 [Bacteriemia coli]	80	64	1200
98	9	6336	7130	gi1467428	YohN [Bacillus subtilis]	80	68	795
98	10	7294	7833	gi1467430	unknown [Bacillus subtilis]	80	64	540
98	11	7820	8737	gi1467431	unknown [Bacillus subtilis]	80	61	918
109	16	14154	14813	gi1580875	high level kasamycin resistance [Bacillus subtilis]	80	61	660
112	15	14294	16636	gi11072361	ipe-57D gene product [Bacillus subtilis]	80	65	2343
139	1	1448	726	gi1506699	pyruvate-formate-lyase [Clostridium pasteurianum]	80	58	723
139	2	2179	1448	gi1506698	CapC [Staphylococcus aureus]	80	59	732
174	4	3271	2870	gi11166242	CapB [Staphylococcus aureus]	80	61	402
					aspartate 1-decarboxylase [Bacillus subtilis]			

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
177	3	2102	2842	gi1467385	unknown [Bacillus subtilis]	80	70	741
184	6	6124	5912	gi1161953	85-kDa surface antigen [Trypanosoma cruzi]	80	46	213
186	4	5368	3875	gi1289282	glutathyl-tryptophan synthetase [Bacillus subtilis]	80	65	1494
205	130	15796	15140	gi140103	ribosomal protein L4 [Bacillus stearothermophilus]	80	66	657
207	1	140	1315	gi1460259	enolase [Bacillus subtilis]	80	67	1176
211	3	1078	1590	gi1410131	ORFX7 [Bacillus subtilis]	80	61	513
235	2	1962	2255	gi1143797	valyl-tryptophan synthetase [Bacillus stearothermophilus] sp[1193] [SVV_BACST VALYL-TRNA SYNTHETASE (EC 6.1.1.9) VALINE--TRNA LIGASE] (VALRS)	80	55	294
239	1	1	1263	gi1143000	proton glutamate symport protein [Bacillus stearothermophilus] p[1526247]S26247 glutamate/aspartate transport protein - Bacillus stearothermophilus	80	59	1263
272	5	2724	2461	gi1708993	hypothetical protein [Bacillus subtilis]	80	54	264
301	3	1446	1111	gi1467418	unknown [Bacillus subtilis]	80	58	336
310	4	5697	4501	gi1117686	lacuC gene product [Staphylococcus xylosus]	80	67	1197
310	6	5258	7006	gi1348053	acetyl-CoA synthetase [Bacillus subtilis]	80	67	1749
310	7	7410	9113	gi1103865	formyl-tetrahydrofolate synthetase [Streptococcus mutans]	80	67	1704
325	3	1114	1389	gi1310325	outer capsid protein [Rotavirus sp.]	80	40	276
337	1	1268	636	gi1537049	ORF_0470 [Escherichia coli]	80	55	633
374	2	939	1228	gi11405448	YneF [Bacillus subtilis]	80	70	300
375	5	3062	3331	gi1467448	unknown [Bacillus subtilis]	80	68	270
388	1	267	587	gi11064791	function unknown [Bacillus subtilis]	90	65	321
394	1	9	459	gi1304976	matches PS00017: ATP-GTP-A and PS00101: EFACITOR GTP: similar to elongation factor G, TetM/tecO tetracycline-resistance proteins [Escherichia coli]	80	65	651
456	1	625	1263	gi11146183	putative [Bacillus subtilis]	80	65	639
475	1	1	654	gi1288269	beta-fructofuranosidase [Staphylococcus xylosus]	80	66	614
544	2	1449	2240	gi1529754	speC [Streptococcus pyogenes]	80	50	792
622	4	1623	1871	gi11483545	unknown [Mycobacterium tuberculosis]	80	65	249
719	1	1	1257	gi11064791	function unknown [Bacillus subtilis]	80	68	1257
739	1	107	838	gi1666983	putative ATP binding subunit [Bacillus subtilis]	80	61	732

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
745	2	581	414	gi1511600	coenzyme PQQ synthetase protein III [Methanococcus jannaschii]	80	61	168
822	1	17	679	gi410141	ORFX17 [Bacillus subtilis]	80	68	661
827	2	901	836	gi1205301	leukotoxin secretion ATP-binding protein [Haemophilus influenzae]	80	54	156
1044	1	3	149	gi160632	pp2 [Marburg virus]	80	55	147
1220	2	571	413	gi1461072	gallidermin precursor - Staphylococcus gallinarum	80	74	159
2519	1	75	275	gi1147556	Opj [Escherichia coli]	80	45	201
2947	1	503	279	gi1184680	polynucleotide phosphorylase [Bacillus subtilis]	80	62	225
3120	1	2	226	gi1517205	67 kDa Myosin-crossreactive streptococcal antigen [Streptococcus yogenes]	80	65	225
3191	1	294	148	gi131259	HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas mvalonii] pir144756 [Pseudomonas sp. hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88)]	80	59	147
3560	2	285	434	gi1217130	photosystem I core protein B [Synecoccus vulcanus]	80	70	150
3655	1	47	346	gi1415855	deoxyribose aldolase [Myoplasma hominis]	80	56	300
3658	2	324	584	gi1551531	2-nitropropene dioxygenase [Miliopsis seturnus]	80	54	261
3769	1	798	400	gi1133950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	80	68	399
3781	1	692	348	gi1166412	NADH-glutamate synthase [Medicago sativa]	80	62	345
3948	1	48	287	gi11204696	fructose-6-phosphate 1,6-bisphosphate [Haemophilus influenzae]	80	69	240
4030	1	571	287	gi11009366	respiratory nitrate reductase [Bacillus subtilis]	80	60	285
4092	1	547	275	gi1170207	orf6 [Lactobacillus sake]	80	69	273
4103	1	680	342	gi139956	lglc [Bacillus subtilis]	80	65	339
4231	1	692	348	gi1289287	UDP-glucose pyrophosphorylase [Bacillus subtilis]	80	65	345
4265	1	595	299	gi1603768	HutI protein, imidazolone-5-propionate hydrolase [Bacillus subtilis] gi1603768 HutI protein, imidazolone-5-propionate hydrolase [Bacillus subtilis]	80	63	297
4504	1	498	250	gi1133950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	80	68	249
2	6	5998	6798	gi1535351	CodY [Bacillus subtilis]	79	63	801
4	7	8295	7057	gi1603768	HutI protein, imidazolone-5-propionate hydrolase [Bacillus subtilis] gi1603768 HutI protein, imidazolone-5-propionate hydrolase [Bacillus subtilis]	79	64	1245
25	6	5273	5515	gi1467281	acyl carrier protein - Rhizobium meliloti	79	65	243

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
59	2	1173	1424	gi1147923	threonine dehydratase 2 [EC 4.2.1.16] [Escherichia coli]	79	75	252
60	1	1	204	gi1666115	orf1 upstream of glucose kinase [Staphylococcus xylosus] pir[S52351][S52351] hypothetical protein 1 - Staphylococcus xylosus	79	60	204
61	1	3002	3590	gi1456882	ppa1: B1496.C2.189 [Mycobacterium leprae]	79	64	1413
65	7	7023	6505	gi1143164	phosphoribosyl aminimidazole carboxylase 1 (PUR-E) [Bacillus subtilis]	79	60	519
49	6	5680	4554	gi1144906	product homologous to E. coli chlorodioxin reductase. J Biol Chem. 1988; 263:9015-9019, and to F52a protein of alkyl hydroperoxide oxidase from S. typhimurium. J Biol Chem. (1990) 265:10535-10540, pen reading frame A [Clostridium pasteurianum]	79	35	1107
102	11	7489	8571	gi1143093	ketol-acid reductoisomerase [Bacillus subtilis] sp[P37253][LUC-BACSU KETOL-ACID REDUCTOISOMERASE (EC 1.1.1.86) ACETOHYDROXY-ACID ISOMEROREDUCTASE] [ALPHA-KETO-BETA-HYDROXYLACIL EDUCOISOMERASE]	79	64	1083
102	14	11190	12563	gi1149428	putative [Lactococcus lactis]	79	65	1374
127	9	7792	9372	gi1458688	PrfC/RF3 [Dichelobacter nodosus]	79	68	1581
139	3	2540	3983	gi1506697	[CapA [Staphylococcus aureus]	79	55	558
144	2	1644	1156	gi11498296	peptide methionine sulfoxide reductase [Streptococcus pneumoniae]	79	47	489
144	2	529	1098	gi1467457	hypoxanthine-guanine phosphoribosyltransferase [Bacillus subtilis] gi1467457 hypoxanthine-guanine phosphoribosyltransferase [Bacillus subtilis]	79	59	570
150	1	965	591	gi1755602	unknown [Bacillus subtilis]	79	61	375
176	1	1039	587	gi1297874	fructose-bisphosphate aldolase [Staphylococcus carnosus] pir[A49943][A49943] fructose-bisphosphate aldolase (EC 4.1.2.13) - taphylococcus carnosus (strain TM300)	79	65	453
186	7	7584	6874	gi11314298	ORF5: putative Sns protein: similar to Sns proteins from Haemophilus influenzae and Escherichia coli [Listeria monocytogenes]	79	64	711
205	16	8887	8498	gi1044980	ribosomal protein L18 [Bacillus subtilis]	79	70	390
211	1	1	519	gi1130394	YohM [Bacillus subtilis]	79	62	519
223	2	4183	2801	gi1488430	alcohol dehydrogenase 2 [Entamoeba histolytica]	79	60	1381
243	8	8915	7896	gi1580883	[ipa-88d gene product [Bacillus subtilis]	79	60	1020
279	4	3721	4329	gi1413930	[ipa-6d gene product [Bacillus subtilis]	79	59	609
300	1	11	1393	gi1403372	[glycerol 3-phosphate permease [Bacillus subtilis]	79	62	1383
307	3	2930	1935	gi1950062	[hypothetical yeast protein 1 [Mycoplasma capricolum] pir[S48578][S48578] [hypothetical protein - Mycoplasma capricolum SOC3] (fragment)]	79	60	996

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
352	6	10106	8886	gi 216854	P47K [Pseudomonas chlororaphis]	79	59	1221
412	1	1153	578	gi 143177	putative [Bacillus subtilis]	79	51	576
481	3	621	1124	gi 786163	Ribosomal Protein L10 [Bacillus subtilis]	79	66	504
516	1	702	352	gi 805090	NI5F [Lactococcus lactis]	79	48	351
525	2	2457	1426	gi 143371	phosphoribosyl aminimidazole synthetase (pur-m) [Bacillus subtilis] pir H29326 AJ85CU phosphoribosylformylglycinamide cyclo-ligase EC 6.3.3.1 - Bacillus subtilis	79	61	1032
538	4	3448	2825	gi 1370207	orf6 [Lactobacillus sake]	79	67	624
570	1	2	421	gi 476160	arginine permease substrate-binding subunit [Listeria monocytogenes]	79	61	420
645	8	2663	3241	gi 153898	transport protein [Salmonella typhimurium]	79	62	579
683	1	75	374	gi 1064795	function unknown [Bacillus subtilis]	79	62	300
816	3	4700	3987	gi 1407784	orf-1; novel antigen [Staphylococcus aureus]	79	62	714
2829	1	3	401	gi 1524397	glycine betaine transporter OpuB [Bacillus subtilis]	79	61	399
2937	1	357	202	pir S52915 S529	nitrate reductase alpha chain - Bacillus subtilis (fragment)	79	58	156
2940	1	768	185	gi 145429	putative [Lactococcus lactis]	79	72	364
2946	1	570	286	gi 143287	2-oxoglutarate dehydrogenase (odhA, EC 1.2.4.2) [Bacillus subtilis]	79	61	285
2979	1	3	212	gi 710020	nitrite reductase (nirB) [Bacillus subtilis]	79	59	210
3022	1	514	332	gi 450686	3-phosphoglycerate kinase [Thermotoga maritima]	79	61	183
3064	1	3	314	gi 1204436	pyruvate formate-lyase [Haemophilus influenzae]	79	60	312
3083	1	2	220	gi 1149662	hnpD gene product [Clostridium perfringens]	79	56	219
3126	1	701	411	gi 1339950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	79	55	291
3181	1	607	326	gi 1339950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	79	59	282
3345	1	3	476	gi 871784	Clp-like ATP-dependent protease binding subunit [Bos taurus]	79	63	474
3718	1	536	270	pir C36889 C368	leuB protein, inactive - Lactococcus lactis subsp. lactis (strain IL1403)	79	71	267
3724	2	159	401	gi 1009366	respiratory nitrate reductase [Bacillus subtilis]	79	64	243
3836	1	608	312	gi 1524193	unknown [Mycobacterium tuberculosis]	79	65	297
3941	1	2	334	gi 415855	deoxyribose aldolase [Mycoplasma hominis]	79	54	333
4113	1	3	341	gi 143015	gluconate kinase [Bacillus subtilis]	79	63	339

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Cunty ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
4501	1	406	209	gi1022726	unknown [Staphylococcus haemolyticus]	79	66	198
4612	1	2	238	gi1460689	TVG [Thermococcus vulgaris]	79	58	237
2	1	2	1213	gi1520753	[DNA topoisomerase I [Bacillus subtilis]	78	66	1212
4	2	2266	1220	gi1216151	[DNA polymerase (gene L; ttg start codon) [Bacteriophage SP02] gi1579197 SP02 DNA polymerase (aa 1-648) [Bacteriophage SP02] p1r[A21498]DJBPS2 DNA- directed DNA polymerase (EC 2.7.7) - phage P02	78	72	1047
9	2	1340	1089	gi1064787	[function unknown [Bacillus subtilis]	78	57	252
32	8	6803	7702	gi1446974	[NH3-dependent NAD synthetase [Escherichia coli]	78	63	900
36	4	2941	3138	gi1290503	[glutamate permease [Escherichia coli]	78	53	198
53	15	117684	16221	gi13303941	[yqiv [Bacillus subtilis]	78	58	1404
57	14	10520	12067	gi11072418	[glcA gene product [Staphylococcus carnosus]	78	65	1548
66	7	6798	5812	gi11212729	[yqhJ [Bacillus subtilis]	78	67	987
67	4	4029	4176	gi1466612	[nika [Escherichia coli]	78	71	148
91	9	10058	10942	gi1467380	[stage 0 sporulation [Bacillus subtilis]	78	50	885
102	12	8574	10130	gi1149426	[putative [lactococcus lactis]	78	61	1557
112	6	3540	4463	gi1854234	[cymC gene product [Klebsiella oxytoca]	78	56	924
124	2	1888	1061	gi1405622	[unknown [Bacillus subtilis]	78	60	878
130	3	1805	2260	gi11256636	[putative [Bacillus subtilis]	78	71	456
133	1	751	377	gi1168060	[lamB [Escherichia nidulans]	78	55	375
166	4	7125	6163	gi1451216	[mannosephosphate isomerase [Streptococcus mutans]	78	63	963
186	1	1586	795	gi1289284	[cysteiny]-tRNA synthetase [Bacillus subtilis]	78	63	792
195	4	2749	2315	gi11351874	[unknown [Rhodospirillum rubrum]	78	58	435
199	3	4279	3623	gi1143525	[succinate dehydrogenase cytochrome b-558 subunit [Bacillus subtilis] p1r[A2984]DESSC succinate dehydrogenase (EC 1.3.99.1) cytochrome 558 - Bacillus subtilis	78	57	657
199	4	7209	5557	gi1142521	[deoxyribodipyrimidine photolyase [Bacillus subtilis] p1r[A37192]A37192 uvrB protein - Bacillus subtilis sp p14951 UVRB_BACSU EXCINUCLEASE ABC SUBUNIT C	78	62	1653
223	3	3831	3523	gi1139596	[Escherichia coli 19200 insertion sequence from ECOR63, partial.), ene product [Escherichia coli]	78	47	309

TABLE 2



S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
299	4	1865	2149	gi1467439	temperature sensitive cell division [Bacillus subtilis]	78	62	285
321	9	7734	7315	gi142979	ORF3 is homologous to an ORF downstream of the spot gene of E.coli; RF3 [Bacillus stearothermophilus]	78	55	420
352	4	3714	3944	gi1349050	actin 1 [Pneumocystis carinii]	78	42	231
352	5	7592	6093	gi1903587	NADH dehydrogenase subunit 5 [Bacillus subtilis] sp P39755 NDHFBACSU NADH DEHYDROGENASE SUBUNIT 5 (EC 1.6.5.3) NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5)	78	58	1500
376	1	2	581	gi1551693	dethiolactin synthase [Bacillus sphaericus]	78	34	582
424	2	1595	1768	gi1524117	alpha-acetolactate decarboxylase [Lactococcus lactis]	78	68	174
450	1	1914	988	gi11010068	NAD(P)H oxidoreductase, isoflavone reductase homologue [Solanum tuberosum]	78	63	927
558	1	762	562	gi1151588	bifunctional protein [Methanococcus jannaschii]	78	60	201
670	3	1152	1589	gi1122759	unknown [Bacillus subtilis]	78	64	438
714	1	64	732	gi1143460	37 kDa minor sigma factor (rpoF, sigB; tly start codon) [Bacillus subtilis]	78	57	689
814	1	3	368	gi11377833	unknown [Bacillus subtilis]	78	59	366
981	1	1381	692	gi1143802	Gerc2 [Bacillus subtilis]	78	64	690
995	2	978	727	gi1296947	uridine kinase [Escherichia coli]	78	64	252
1045	1	3	801	gi11407784	orf-1; novel antigen [Staphylococcus aureus]	78	61	399
1161	2	168	186	gi1410117	diaminopimelate decarboxylase [Bacillus subtilis]	78	54	183
2191	1	794	399	gi1215098	lecithinase [Bacteriophage 154a]	78	65	396
2933	1	2	181	gi11204636	pyruvate formate-lyase [Haemophilus influenzae]	78	73	180
3041	2	129	317	gi1624632	GltC [Escherichia coli]	78	53	189
3581	1	105	401	gi1763186	3-ketoacyl-CoA thiolase [Saccharomyces cerevisiae]	78	55	237
3709	1	3	230	gi1460689	TWg [Thermactinomyces vulgaris]	78	58	218
3974	1	528	265	gi1558839	unknown [Bacillus subtilis]	78	65	264
3980	1	3	401	gi139956	ITGlc [Bacillus subtilis]	78	62	399
4056	1	647	354	gi1256635	aldehyde-acyl dehydrogenase [Bacillus subtilis]	78	55	294
4114	1	630	316	gi1509372 5093	hypothetical protein - Trypanosoma brucei	78	62	315
4185	1	3	179	gi11339950	large subunit of NADH-dependent glutamate synthase [Plectononema boryanum]	78	58	177

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
4215	1	655	109	gi 558819	unknown [Bacillus subtilis]	78	60	327
4352	1	541	302	gi 603768	MutI protein, imidazolone-5-propionate hydrolase [Bacillus subtilis] gi 603768 MutI protein, imidazolone-5-propionate hydrolase [Bacillus subtilis]	78	63	240
4368	1	612	307	gi 1153678	heavy-metal transporting P-type ATPase [Proteus mirabilis]	78	59	306
4461	1	428	216	gi 1276841	glutamate synthase (GOGAT) [Porphyra purpurea]	78	36	213
4530	1	474	238	gi 139956	11Glc [Bacillus subtilis]	78	65	237
3	2	2969	2073	gi 1109684	ProV [Bacillus subtilis]	77	56	897
12	2	2426	1965	gi 467335	ribosomal protein L9 [Bacillus subtilis]	77	59	462
27	1	2	388	gi 1212728	YqhI [Bacillus subtilis]	77	63	347
39	2	590	1252	gi 40054	phenylalanyl-tRNA synthetase beta subunit (AA 1-804) [Bacillus subtilis]	77	60	663
42	6	2704	2931	gi 608241	30S ribosome subunit protein S14 [Escherichia coli ap P02370 RS14_ECOLI 30S RIBOSOMAL PROTEIN S14 (SUB 2-101)]	77	65	228
46	18	15459	16622	gi 297798	mitochondrial formate dehydrogenase precursor [Solanum tuberosum] pif J02272 J02572 formate dehydrogenase (EC 1.2.1.2) precursor, mitochondrial - potato	77	55	1164
100	4	4562	4002	gi 1340128	ORF1 [Staphylococcus aureus]	77	54	561
102	8	5378	5713	gi 1311482	acetolactate synthase [Thermus aquaticus]	77	57	336
109	7	4742	5383	gi 710637	unknown [Bacillus subtilis]	77	56	642
117	1	2	1228	gi 1237015	ORF4 [Bacillus subtilis]	77	53	1227
124	10	8323	7688	gi 405819	thymidine kinase [Bacillus subtilis]	77	63	636
147	3	1146	985	gi 849027	hypothetical 15.9-kDa protein [Bacillus subtilis]	77	37	162
152	10	7354	7953	gi 1205583	spermidine/putrescine transport ATP-binding protein [Haemophilus influenzae]	77	55	600
169	2	1004	1282	gi 473825	'elongation factor EF-Ts' [Escherichia coli]	77	58	279
184	2	180	1147	gi 216314	esterase [Bacillus stearothermophilus]	77	60	768
189	7	3296	3868	gi 853809	ORF3 [Clostridium perfringens]	77	48	573
193	1	132	290	gi 1303788	YqeH [Bacillus subtilis]	77	54	159
195	8	8740	8414	gi 1499620	M. jannaschii predicted coding region M30798 [Methanococcus jannaschii]	77	44	327
205	8	5428	5204	gi 216340	ORF for adenylate kinase [Bacillus subtilis]	77	61	225

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match	match gene name	% sim	% ident	length (nt)
205	29	14795	14502	gi1786155	gi1786155	Ribosomal Protein L23 [Bacillus subtilis]	77	62	294
211	5	1908	2084	gi1410132	gi1410132	OMF8 [Bacillus subtilis]	77	47	177
217	5	3478	4416	gi1496254	gi1496254	fibrinectin/fibrinogen-binding protein [Streptococcus pyogenes]	77	54	939
232	1	267	998	gi1407784	gi1407784	orf-1; novel antigen [Staphylococcus aureus]	77	57	732
233	2	1819	1346	gi1467408	gi1467408	unknown [Bacillus subtilis]	77	61	474
243	3	2661	2299	gi1516155	gi1516155	unconventional myosin [Sus scrofa]	77	32	363
249	1	6A	769	gi1467436	gi1467436	unknown [Bacillus subtilis]	77	54	702
301	4	1468	1283	gi1950071	gi1950071	ATP-bind. pyrimidine kinase [Mycoplasma capricolum] pir[S46605]S46605	77	48	186
302	5	2741	3211	gi1508980	gi1508980	hypothetical protein - Mycoplasma capricolum SGC3 (fragment)	77	57	471
302	7	3835	4863	gi147783	gi147783	trpB protein [Escherichia coli]	77	60	1029
307	9	5402	4797	gi1070015	gi1070015	protein-dependent [Bacillus subtilis]	77	60	606
312	1	99	1391	gi143165	gi143165	malic enzyme (EC 1.1.1.38) [Bacillus stearothermophilus] pir[A33307]DEBSXS	77	62	1293
						malate dehydrogenase oxaloacetate-decarboxylating (EC 1.1.1.38) - Bacillus stearothermophilus			
312	2	1541	2443	gi1398855	gi1398855	carboxyltransferase beta subunit [Synecococcus RCC7942]	77	58	903
321	5	5666	4596	gi139844	gi139844	funarase (citG) (aa 1-462) [Bacillus subtilis]	77	65	1071
354	1	47	568	gi1154634	gi1154634	YmoR [Bacillus subtilis]	77	57	522
365	1	2	1021	gi143374	gi143374	phosphoribosyl glycine synthetase (PUR-D; gta start codon) Bacillus subtilis	77	62	1020
374	1	1	708	gi1405446	gi1405446	transketolase [Bacillus subtilis]	77	61	708
385	1	1128	565	gi1533099	gi1533099	endonuclease III [Bacillus subtilis]	77	63	564
392	2	594	1940	gi1556014	gi1556014	UDP-N-acetyl muramate-alanine ligase [Bacillus subtilis] sp[910778]MURC-BACSU UDP-N-ACETYLURAMATE--ALANINE LIGASE (EC 3.2.8)	77	65	1347
						UDP-N-ACETYLURAMATE--ALANINE SYNTHETASE (FRAGMENT)			
405	5	4079	3570	gi11303912	gi11303912	Yqmw [Bacillus subtilis]	77	64	510
487	4	1302	1472	gi1432437	gi1432437	ORF1 gene product [Acinetobacter calcoaceticus]	77	48	171
522	1	2	562	gi1401179	gi1401179	tyrosine-tRNA ligase (EC 6.1.1.1) - Bacillus stearothermophilus	77	63	561

TABLE 2

S. aureus Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Myth gene name	% sim	% ident	length (nt)
521	2	1547	1351	gi11387979	44% identity over 302 residues with hypothetical protein from <i>Synechocystis</i> sp. accession D64006.CD; expression induced by environmental stress, some similarity to glycosyl transferases, two potential membrane spanning helices [Bacillus subtilis]	77	48	237
536	2	983	612	gi1140366	adenylosuccinate lyase [Pur-B] [Bacillus subtilis] pir[C29326]w2B505	77	61	372
548	2	339	872	gi1140387	adenylosuccinate lyase (EC 4.3.2.2) - Bacillus subtilis	77	56	514
597	1	2	481	gi1904198	aspartate transcarbamylase [Bacillus subtilis]	77	33	480
631	2	1747	1313	gi1387577	hypothetical protein [Bacillus subtilis]	77	64	435
642	1	85	360	gi146971	ORF1A [Bacillus subtilis]	77	61	276
659	1	125	1219	gi11072381	epip gene product [Staphylococcus epidermidis]	77	62	1095
670	4	1587	1820	gi1122760	glutamy-aminopeptidase [Lactococcus lactis]	77	58	234
789	1	2	391	gi11377823	unknown [Bacillus subtilis]	77	65	390
815	1	10	573	gi11303861	aminopeptidase [Bacillus subtilis]	77	49	564
890	1	1	225	gi11101444	YqgH [Bacillus subtilis]	77	55	225
1083	1	3	188	gi1460828	III, Influenzan predicted coding region H10594 [Haemophilus influenzae]	77	66	186
1942	1	415	209	gi1160047	10K malarial antigen precursor - Plasmodium alciptarum (strain Camp)	77	38	207
2559	1	1	171	gi11499034	pl01/acidic basic repeat antigen [Plasmodium falciparum] pir[A29232]A29232	77	61	171
2933	2	24	401	gi142370	M. jamaensis predicted coding region M2055 [Methanococcus jamaensis]	77	72	159
2966	1	56	292	gi11524397	pyruvate formate-lyase (AA 1-760) [Escherichia coli] irf1501788[S01788 formate C-acetyltransferase (EC 2.3.1.54) - Escherichia coli]	77	45	237
2976	1	614	309	gi140003	glycine betaine transporter Opud [Bacillus subtilis]	77	60	306
2979	2	678	400	gi1204354	oxoglutarate dehydrogenase (NADP+) [Bacillus subtilis] p123129[ODO]-BACSU 2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT (EC 2.4.2) (ALPHA-KETOGLUTARATE DEHYDROGENASE)	77		
2984	1	601	377	gi1438465	spore germination and vegetative growth protein [Haemophilus influenzae]	77	61	279
2990	1	331	167	gi1142562	Probable operon with orf1 Possible alternative initiation codon, aaaa 2151-2153. Homology with acetyltransferases; putative Bacillus subtilis	77	55	225
3032	1	3	389	gi1488430	ATP synthase epsilon subunit [Bacillus megaterium] pir[B28599]PMBSEN H-transferring ATP synthase (EC 3.6.1.34) psilon chain - Bacillus megaterium	77	63	165
3057	1	1	195	gi1468764	alcohol dehydrogenase 2 [Entamoeba histolytica]	77	56	387
					moct gene product [Rhizobium meliloti]	77	50	195

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match (%)	Gene name	% sim	% ident	length (nt)
4008	1	726	400	gi 603768		HutI protein, imidazole-5-propionate hydrolase [Bacillus subtilis] gi 603768 HutI protein, imidazole-5-propionate hydrolase Bacillus subtilis	77	52	327
4248	1	703	386	gi 216278		Gramicidin S synthetase 1 [Bacillus brevis]	77	55	318
4110	1	3	388	pir 552915 5529		Nitrate reductase alpha chain - Bacillus subtilis (fragment)	77	61	366
4115	1	1	348	gi 517205		67 kDa Myosin-crossreactive streptococcal antigen [Streptococcus yogenes]	77	65	348
4225	1	590	297	gi 1322245		Nevalonate pyrophosphate decarboxylase [Rattus norvegicus]	77	60	294
4611	2	494	327	gi 508879		GTP-binding protein [Bacillus subtilis]	77	57	188
466A	1	361	182	pir 552915 5529		Nitrate reductase alpha chain - Bacillus subtilis (fragment)	77	61	180
25	1	2	1627	gi 1150620		HnsA [Streptococcus pneumoniae]	76	58	1626
3A	5	1488	2537	pir A43577 A435		Regulatory protein pfor - Clostridium perfringens	76	57	1050
52	5	2962	4041	gi 1161061		Dioxygenase [Methylobacterium extorquens]	76	62	1080
56	120	12789	27955	gi 467402		Unknown [Bacillus subtilis]	76	56	567
57	15	12046	12219	gi 1206040		Weak similarity to keratin [Caenorhabditis elegans]	76	40	174
91	2	1062	2261	gi 475715		Acetyl coenzyme A acetyltransferase (thiolase) [Clostridium acetobutylicum]	76	57	1200
9A	2	818	1624	gi 467422		Unknown [Bacillus subtilis]	76	62	807
9A	5	2965	3228	gi 897793		Y98 gene product [Pediococcus acidilactici]	76	52	264
9B	8	5922	6326	gi 467427		Methionyl-tRNA synthetase [Bacillus subtilis]	76	51	405
104	3	1322	1885	gi 216151		tRNA polymerase (gene U; ttc start codon) [Bacteriophage SP02] gi 579197 SP02 DNA polymerase (aa 1-648) [Bacteriophage SP02] pir A21498 Q21498 DNA-directed DNA polymerase (EC 2.7.7.7) - phage P02	76	63	564
124	9	8134	7055	gi 853776		Peptide chain release factor 1 [Bacillus subtilis] pir 555437 555437 peptide chain release factor 1 - Bacillus subtilis	76	58	1080
164	5	2832	3311	gi 1204976		Prolyl-tRNA synthetase [Haemophilus influenzae]	76	51	480
168	2	2617	1841	gi 1177253		Putative ATP-binding protein of ABC-type [Bacillus subtilis]	76	58	777
189	2	163	888	gi 467384		Unknown [Bacillus subtilis]	76	63	726
235	3	2253	3518	gi 142936		Polyl-polyglutamate synthetase [Bacillus subtilis] pir 840646 840646 foLC - Bacillus subtilis	76	51	1266
236	1	335	925	gi 1146197		Putative [Bacillus subtilis]	76	54	591
237	8	5323	5541	gi 1279261		F13G3.6 [Caenorhabditis elegans]	76	47	219

TABLE 2

S. aureus - putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
263	5	5490	4585	gi11510148	[dihydrodipicolinate synthase [Methanococcus jannaschii]]	76	49	906
304	3	1051	1794	gi1666982	[putative membrane spanning subunit [Bacillus subtilis] pir S52382 S52382 probable membrane spanning protein - Bacillus subtilis]	76	60	744
312	4	3611	4624	gi1143312	[6-phospho-1-fructokinase [gtg start codon: EC 2.7.1.11] [Bacillus therophilus]]	76	56	1014
343	1	2	1016	gi1405956	[yeeE [Escherichia coli]]	76	59	1035
347	1	409	1701	gi1396304	[acetylornithine decarboxylase [Escherichia coli]]	76	72	1293
358	1	672	1907	gi1146215	[39.0% identity to the Escherichia coli S1 ribosomal protein; putative [Bacillus subtilis]]	76	58	1236
371	1	1	222	gi1537084	[alternate gene name mgt: CG Site No. 497 [Escherichia coli] pir S56468 S56468 mgtA protein - Escherichia coli]	76	61	222
379	4	4331	4858	gi1143268	[dihydrolipoamide transsuccinylase (odhB; EC 2.3.1.61) [Bacillus subtilis]]	76	61	528
404	5	4022	4492	gi1330823	[YqfC [Bacillus subtilis]]	76	60	471
413	1	2	307	gi1386025	[ORF YKL027W [Saccharomyces cerevisiae]]	76	55	306
472	3	4356	2854	gi11405464	[Alat [Bacillus subtilis]]	76	57	1503
546	1	273	995	gi1153821	[streptococcal pyrogenic exotoxin type C (speC) precursor Streptococcus pyogenes]	76	36	723
588	1	1054	557	gi11002520	[MucS [Bacillus subtilis]]	76	61	498
591	1	16	735	gi1885934	[ClpB [Synchococcus sp.]	76	44	720
602	2	175	798	gi11486422	[OppD homologue [Rhizobium sp.]	76	52	624
619	2	547	290	gi1330613	[major capsid protein [Human cytomegalovirus]]	76	47	258
660	4	2568	3302	gi1904199	[hypothetical protein [Bacillus subtilis]]	76	55	735
677	1	452	228	gi140177	[spoOF gene product [Bacillus subtilis]]	76	58	225
962	1	24	206	gi1142443	[adenylosuccinate synthetase [Bacillus subtilis] sp p29726 PURA_BACSU ADENYLOSUCCINATE SYNTHETASE (EC 6.3.4.4) IMP-ASPARTATE LIGASE]	76	67	183
978	1	1158	580	gi11511333	[M. jannaschii predicted coding region Mj1322 [Methanococcus jannaschii]]	76	56	579
997	1	486	244	gi1467154	[no definition line found [Mycobacterium leprae]]	76	38	243
1563	1	529	266	gi13303984	[YqkG [Bacillus subtilis]]	76	52	264
2184	1	361	182	gi1506706	[CapJ [Staphylococcus aureus]]	76	38	180
2572	1	1	387	gi1153898	[transport protein [Salmonella typhimurium]]	76	65	387

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
2942	1	29	400	gi 710020	nitrite reductase (nirB) [Bacillus subtilis]	76	69	372
2957	1	377	216	gi 1511251	hypothetical protein (SP-P42404) [Methanococcus jannaschii]	76	47	162
2980	1	554	279	gi 1405484	AlaT [Bacillus subtilis]	76	53	276
3015	1	649	326	gi 408115	ornithine acetyltransferase [Bacillus subtilis]	76	61	324
3124	1	13	174	gi 882705	ORF_0401 [Escherichia coli]	76	65	162
3179	1	3	161	gi 168477	ferredoxin-dependent glutamate synthase (Zea mays) pir[AJ8596]AJ8596 glutamate synthase (ferredoxin) (EC 1.4.7.1) - alae	76	53	158
3789	1	2	379	gi 39956	IIGlc [Bacillus subtilis]	76	55	378
3892	1	3	314	gi 1510198	ferritinocelin binding protein [Methanococcus jannaschii]	76	52	312
3928	1	798	400	gi 143016	permease [Bacillus subtilis]	76	59	398
4159	1	757	386	sp P80544 MBSP_1	METHICILLIN-RESISTANT SURFACE PROTEIN (FRAGMENTS)	76	66	372
4204	1	17	331	gi 296464	ATPase [Lactococcus lactis]	76	56	315
4394	1	494	249	gi 987255	Menkes disease gene (Homo sapiens)	76	48	246
4506	1	2	313	gi 216746	D-lactate dehydrogenase [Lactobacillus plantarum]	76	47	312
4546	1	477	247	gi 133950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	76	61	231
4596	1	379	191	gi 560027	cellulose synthase [Acetobacter xylinum]	76	70	189
4	5	5257	4337	gi 882532	ORF_0294 [Escherichia coli]	75	59	921
6	1	164	952	gi 40960	OTCase [Escherichia coli]	75	56	789
12	3	5935	3944	gi 467336	unknown [Bacillus subtilis]	75	57	1992
23	18	18272	17310	gi 1296433	O-acetylserine sulfinhydrilase B [Alcaligenes eutrophus]	75	55	963
25	3	2356	3393	gi 1502419	Plax [Bacillus subtilis]	75	56	1038
36	8	5765	6037	gi 1256517	unknown [Schizosaccharomyces pombe]	75	45	233
46	13	11186	12058	gi 48972	Nitrate transporter [Synechococcus sp.]	75	46	873
51	7	3474	3677	gi 113607	sporulation protein [Bacillus subtilis]	75	61	204
53	16	16850	16590	gi 143402	recombination protein (ttg start codon) [Bacillus subtilis] gi 1303923 RecN [Bacillus subtilis]	75	51	261
74	3	3572	2568	gi 1204847	ornithine carbamoyltransferase [Haemophilus influenzae]	75	61	1005

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
R5	3	4628	3910	gi146368	phosphoribosylglycinamide synthetase I (pur-L; gta start odon) (Bacillus subtilis)	75	63	699
R5	5	5588	4878	gi146367	phosphoribosyl aminodazole succinocarboxamide synthetase (pur-C; tga start codon) (Bacillus subtilis)	75	55	711
R5	8	6625	7530	gi1303916	Yqia (Bacillus subtilis)	75	53	906
R7	3	2340	3590	gi1064813	homologous to sp:PHO_BACSU (Bacillus subtilis)	75	56	1251
R7	6	6084	6896	gi1064810	function unknown (Bacillus subtilis)	75	61	813
R8	2	1844	1503	gi1001824	hypothetical protein (Synecocystis sp.)	75	51	342
R10	3	1748	3727	gi1147593	putative pGnp synthetase (Streptomyces coelicolor)	75	55	1980
R10	7	4151	5452	gi1117251	ciwD gene product (Bacillus subtilis)	75	75	900
R20	14	11266	10649	gi1524394	ORF-2 upstream of gsaB operon (Bacillus subtilis)	75	55	618
R21	5	2050	4221	gi1154632	Nrde (Bacillus subtilis)	75	54	2172
R24	1	283	143	gi1405622	unknown (Bacillus subtilis)	75	56	141
R28	1	81	1139	gi143316	[gnp] gene products (Bacillus megaterium)	75	48	1059
R30	8	5760	5903	gi1256654	54.8% identity with Neisseria gonorrhoeae regulatory protein pilB; putative (Bacillus subtilis)	75	62	144
R36	2	4480	3185	gi1467403	seryl-tRNA synthetase (Bacillus subtilis)	75	54	1296
R61	10	5439	5798	gi1001195	hypothetical protein (Synecocystis sp.)	75	55	360
R72	4	3819	2995	gi1755153	ATP-binding protein (Bacillus subtilis)	75	52	825
R79	1	2024	1107	gi143037	porphobilinogen deaminase (Bacillus subtilis)	75	58	918
R95	10	9539	9174	gi1257451	HYPOTHETICAL PROTEIN IN PURB 5' REGION (ORF-15) (FRAGMENT)	75	60	156
R200	4	2605	4596	gi142440	ATP-dependent nuclease (Bacillus subtilis)	75	56	1992
R206	3	6900	5620	gi1256135	YdbP (Bacillus subtilis)	75	53	1281
R216	2	159	389	gi1052800	unknown (Schizosaccharomyces pombe)	75	58	231
R229	1	29	847	gi1205958	branched chain aa transport system II carrier protein (Haemophilus influenzae)	75	49	819
R230	2	518	1714	gi1971337	nitrite extrusion protein (Bacillus subtilis)	75	53	1197
R231	1	2240	1122	gi1002521	MutL (Bacillus subtilis)	75	54	1119
R233	3	1314	1859	gi1467405	unknown (Bacillus subtilis)	75	59	546

TABLE 2



S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
268	1	325	164	[gi1511246]	methyl coenzyme M reductase system, component A2 [Methanococcus jannaschii]	75	50	162
292	1	1389	172	[gi1511604]	M. jannaschii predicted coding region M1651 [Methanococcus jannaschii]	75	46	614
304	4	1773	2261	[gi1205328]	surfactin [Methophila influenzae]	75	55	489
312	3	2437	3387	[gi1285621]	undefined open reading frame [Bacillus stearothermophilus]	75	62	951
312	5	4622	6403	[gi1041097]	Pyruvate Kinase [Bacillus psychrophilus]	75	57	1782
319	1	353	877	[gi1121278]	Yqhi [Bacillus subtilis]	75	54	525
320	5	4321	5031	[gi11070361]	OMP decarboxylase [Lactococcus lactis]	75	56	711
320	6	5010	5642	[gi1143394]	OMP-PRPP transferase [Bacillus subtilis]	75	60	633
337	4	1519	2044	[gi1487433]	citrate synthase II [Bacillus subtilis]	75	58	570
394	2	669	1271	[gi1304976]	matches PS00017: ATP-GTP-A and PS00301: EFATOR-GTP, similar to elongation factor G, TetM/TetO (tetracycline-resistance proteins Echerichia coli)	75	51	603
423	1	327	570	[gi11183939]	unknown [Pseudomonas aeruginosa]	75	59	444
433	2	1603	1929	[gi1149211]	lactate synthase [Klebsiella pneumoniae]	75	63	327
436	2	176	1540	[gi1112441]	ldihydroxotetase [Bacillus caldolyticus]	75	62	1365
486	1	494	249	[gi11145682]	potF gene product [Clostridium perfringens]	75	55	246
496	1	3	794	[gi1143582]	IsolHCA protein [Bacillus subtilis]	75	59	792
498	2	824	1504	[gi1143124]	phoP protein (put.): putative [Bacillus subtilis]	75	47	681
499	2	1061	1624	[gi11387979]	44% identity over 302 residues with hypothetical protein from Synchocystis sp. accession D54006_CD; expression induced by environmental stress, some similarity to glycosyl transferases; two potential membrane-spanning helices [Bacillus subtilis]	75	51	564
568	1	841	453	[gi104110341]	triacylglycerol lipase (EC 3.1.1.3) 2 - Mycoplasma mycoides subsp. mycoides [ISGC3]	75	50	189
613	2	430	233	[gi1330993]	tegument protein [Saimiriina herpesvirus 2]	75	75	198
621	1	1	525	[gi1529754]	IspeC [Streptococcus pyogenes]	75	43	525
642	5	1809	2474	[gi1176403]	IsnG [Staphylococcus epidermidis]	75	51	666
646	2	454	657	[gi1172442]	ribonuclease P [Saccharomyces cerevisiae]	75	37	204
657	1	3	347	[gi1802561]	IOUF_0236 [Echerichia coli]	75	47	345
750	1	1662	832	[gi146971]	lepiP gene product [Staphylococcus epidermidis]	75	57	831

TABLE 2

S. aureus Purative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
754	1	2	481	gi 1303901	YqHT [Bacillus subtilis]	75	57	480
763	2	563	393	gi 3205145	multidrug resistance protein [Haemophilus influenzae]	75	51	171
775	1	961	482	pir B36889 B368	leuA protein, inactive - Lactococcus lactis subsp. lactis (strain IL1403)	75	63	480
793	1	1	180	gi 143316	[gap] gene products [Bacillus megaterium]	75	57	180
800	1	318	160	gi 569411	NPRA protein [Azorhizobium caulinodans]	75	34	159
811	1	1117	560	gi 143434	Rho factor [Bacillus subtilis]	75	60	558
940	1	493	329	gi 1276985	arginase [Bacillus caldovelox]	75	50	165
971	2	37	252	gi 1001373	hypothetical protein [Synecocystis sp.]	75	58	216
1059	1	384	232	gi 726480	L-glutamine-D-fructose-6-phosphate amidotransferase [Bacillus subtilis]	75	67	153
1109	2	219	374	gi 143331	alkaline phosphatase regulatory protein [Bacillus subtilis] pir A27650 A27650 regulatory protein phos - Bacillus subtilis sp P23545 PHOS_BACSU ALKALINE PHOSPHATASE SYNTHESIS SENSOR PROTEIN PHOS (EC 2.7.3.1)	75	53	156
1268	1	271	137	gi 304135	ornithine acetyltransferase [Bacillus stearothermophilus] sp Q07908 ARGO_BACST GLUTAMATE N-ACETYLTRANSFERASE (EC 2.3.1.35) ORNITHINE ACETYLTRANSFERASE / ORNITHINE TRANSACETYLASE / IDATASE / HING-ACID ACETYLTANSFRASE (EC 2.3.1.1) (N-ACETYLGUTAMATE YNTIA	75	63	135
1500	1	324	163	gi 1205488	lexinuclease ABC subunit B [Haemophilus influenzae]	75	57	162
1529	1	798	400	gi 1002521	MutL [Bacillus subtilis]	75	54	399
3010	1	720	387	gi 1204435	pyruvate formate-lyase activating enzyme [Haemophilus influenzae]	75	54	384
3105	1	1	180	gi 1041097	Pyruvate Kinase [Bacillus psychrophilus]	75	57	180
3117	1	45	212	gi 899317	peptide synthetase module [Microcystis aeruginosa] pir S4911 S4911 probable amino acid activating domain - Microcystis aeruginosa (fragment) [SUB 144-528]	75	42	168
3139	2	139	345	gi 145294	adenine phosphoribosyl-transferase [Escherichia coli]	75	66	207
3880	1	618	310	gi 1009366	respiratory nitrate reductase [Bacillus subtilis]	75	58	309
3911	1	48	401	gi 433991	ATP synthase subunit beta [Bacillus subtilis]	75	68	354
3957	1	2	379	pir B16885 B168	3-isopropylmalate dehydratase (EC 4.2.1.33) chain leuC - Lactococcus lactis subsp. lactis (strain IL1403)	75	65	378
4005	1	5	259	gi 216746	D-lactate dehydrogenase [Lactobacillus plantarum]	75	48	255
4080	1	73	333	gi 415855	deoxyribose aldolase [Mycoplasma hominis]	75	59	261

TABLE 2

TABLE 2

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4131	1	1	339	gi 149435	putative [Lactococcus lactis]	75	57	339
4136	1	602	303	gi 450688	hsm gene of Ecopit1 gene product [Escherichia coli] pir S38437 S38437 hsm protein - Escherichia coli pir S09629 S09629 hypothetical protein A - Escherichia coli (S08 40-520)	75	56	300
4144	1	668	336	gi 148972	nitrate transporter [Synechococcus sp.]	75	49	333
4237	1	664	374	gi 1339950	large subunit of NADH dependent glutamate synthase [Plectonema boryanum]	75	55	291
4306	2	73	318	gi 1294260	major surface glycoprotein [Pneumocystis carinii]	75	64	242
4343	1	715	359	gi 1204652	methylated-DNA - protein-cysteine methyltransferase [Haemophilus influenzae]	75	52	353
4552	1	620	312	gi 1296464	ATPase [Lactococcus lactis]	75	55	309
48	9	5376	6126	gi 1443793	[NupC [Escherichia coli]	74	50	351
50	8	6910	6221	gi 1239988	hypothetical protein [Bacillus subtilis]	74	55	690
56	9	10770	12221	gi 1000451	[Trep [Bacillus subtilis]	74	57	1452
64	2	2766	1622	gi 141015	arabidate-tRNA ligase [Escherichia coli]	74	57	645
66	6	5063	4848	gi 1212729	[YohB [Bacillus subtilis]	74	47	216
67	18	14334	14897	gi 1510631	endoglucanase [Mechanococcus jannaschii]	74	52	564
102	15	12561	13116	gi 149429	putative [Lactococcus lactis]	74	67	576
102	16	13121	14419	gi 149435	putative [Lactococcus lactis]	74	57	1299
108	4	4873	3902	gi 139478	ATP binding protein of transport ATPase [Bacillus firmus] tr S15486 S15486 ATP-binding protein - Bacillus firmus p P28946 P28946 HYPOTHETICAL ATP-BINDING TRANSPORT PROTEIN	74	59	912
116	5	8574	7093	gi 1205430	dipeptide transport system permease protein [Haemophilus influenzae]	74	49	1482
120	7	4142	4803	gi 146970	ribonucleoside triphosphate reductase [Escherichia coli] pir A47331 A47331 anaerobic ribonucleotide reductase - Escherichia coli	74	58	462
121	7	5961	6581	gi 1107528	ttg start [Campylobacter coli]	74	51	621
128	3	2320	3531	gi 140318	phosphoglycerate kinase [Bacillus megaterium]	74	57	1212
130	7	5237	5791	gi 1256653	DNA-binding protein [Bacillus subtilis]	74	60	555
136	3	6745	5150	gi 143076	histidase [Bacillus subtilis]	74	58	1596
145	2	644	1368	gi 1407773	devA gene product [Anabaena sp.]	74	45	105
152	1	552	277	gi 1377833	unknown [Bacillus subtilis]	74	54	276

S. aureus Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	saafp gene name	% sim	% ident	length (nt)
164	110	11064	11375	gi1580900	ORF3 gene product [Bacillus subtilis]	74	52	312
175	2	3109	2624	gi1642656	unknown [Rhizobium meliloti]	74	34	466
175	9	8064	5612	gi1854656	Na/H antiporter system ORF2 [Bacillus alcalophilus]	74	46	453
195	11	11346	10339	gi11204430	hypothetical protein (SP-P25745) [Haemophilus influenzae]	74	55	1008
205	17	9619	9059	gi11044879	ribosomal protein L6 [Bacillus subtilis]	74	64	561
236	7	5574	6710	gi11146207	putative [Bacillus subtilis]	74	63	1137
241	3	4521	3336	gi1694121	malate thiokinase [Methylobacterium extorquens]	74	52	1188
246	6	3305	2799	gi1467374	single strand DNA binding protein [Bacillus subtilis]	74	64	507
249	4	6551	5313	gi11524397	glycine betaine transporter OpuD [Bacillus subtilis]	74	55	1239
261	7	4389	4081	gi1809542	CbrB protein [Erwinia chrysanthemi]	74	42	309
278	6	5714	4665	gi11204872	ATP-binding protein [Haemophilus influenzae]	74	54	1050
309	1	1220	666	gi11205579	hypothetical protein (CB-U14003.302) [Haemophilus influenzae]	74	53	555
315	2	1473	862	gi1143398	quinol oxidase [Bacillus subtilis]	74	57	612
320	1	1	1065	gi1343389	glutaminase of carbamyl phosphate synthetase [Bacillus subtilis] pir[E39845]E39845 carbamoyl-phosphate synthase glutamine-hydrolyzing (EC 6.3.5.5), pyrimidine-repressible, small chain - Bacillus subtilis	74	60	1065
380	2	382	1128	gi1534857	ATPase subunit a [Bacillus stearothermophilus]	74	56	747
405	2	1742	1311	gi11303915	YqhZ [Bacillus subtilis]	74	65	432
433	5	2503	3270	gi1473902	alpha-acetolactate synthase [Bacillus subtilis]	74	56	768
452	1	1	942	gi1413982	lpa-58r gene product [Bacillus subtilis]	74	52	942
461	1	3	1193	gi1558494	homoserine dehydrogenase [Bacillus subtilis]	74	51	1191
461	2	1174	1407	gi140211	threonine synthase (thc) (AA 1-352) [Bacillus subtilis] ir[A25364]A25364 threonine synthase (EC 4.2.99.2) - Bacillus subtilis	74	56	234
462	2	402	734	gi142520	thioredoxin [Bacillus subtilis]	74	62	333
478	1	574	320	gi1469005	glycyl-tRNA synthetase [Methanococcus jannaschii]	74	52	255
501	2	739	1740	gi1217040	acid glycoprotein [Streptococcus pyogenes]	74	58	1002
551	2	4083	2791	gi143040	glutamate-L-semialdehyde 2,1-aminotransferase [Bacillus subtilis] pir[D42728]D42728 glutamate-L-semialdehyde 2,1-aminotransferase (EC 4.3.8.8) - Bacillus subtilis	74	51	1293

TABLE 2

S. aureus - putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
573	1	1	477	gi11006605	hypothetical protein [Synecocystis sp.]	74	45	477
596	2	1780	1298	gi11301853	YqgF [Bacillus subtilis]	74	55	483
618	2	2924	1758	gi11146237	21.4% of identity to trans-acting transcription factor of <i>Saccharomyces cerevisiae</i> ; 25% of identity to sucrose synthase of <i>Zea mays</i> ; putative [Bacillus subtilis]	74	55	1167
659	2	1269	1595	gi11072380	ORF3 [Lactococcus lactis]	74	62	327
724	1	373	188	gi1143374	(phosphoribosyl) glycinamide synthetase (Pur-D; gta start codon) [Bacillus subtilis]	74	58	186
743	2	604	1209	gi1153833	ORF1; putative [Streptococcus parasanguis]	74	50	606
836	1	2	259	gi1143458	ORF V [Bacillus subtilis]	74	47	258
989	2	443	724	gi11303994	YqgM [Bacillus subtilis]	74	46	282
1106	1	1	492	gi1146970	epiD gene product [Staphylococcus epidermidis]	74	54	492
1135	2	373	528	gi11413948	ipe-24d gene product [Bacillus subtilis]	74	48	156
1234	1	817	432	gi1495245	recJ gene product [Erwinia chrysanthemi]	74	36	360
2586	1	2	238	gi11149701	sbcC gene product [Clostridium perfringens]	74	62	237
2959	1	798	400	gi11405454	aconitase [Bacillus subtilis]	74	60	399
2962	1	650	163	gi1450886	3-phosphoglycerate kinase [Thermotoga maritima]	74	58	288
2983	1	3	191	gi11303893	YqgH [Bacillus subtilis]	74	56	189
3018	1	2	223	gi143040	glutamate-2-semialdehyde 2,1-aminotransferase [Bacillus subtilis] pir10427281042728 glutamate-2-semialdehyde 2,1-aminotransferase (EC 4.3.8) - [Bacillus subtilis]	74	56	222
3038	1	510	246	pir155291515529	nitrate reductase alpha chain - [Bacillus subtilis (fragment)]	74	57	255
3062	1	374	189	gi1107528	ftsG start [Campylobacter coli]	74	51	186
4035	1	184	360	gi11022725	unknown [Staphylococcus haemolyticus]	74	64	177
4045	1	607	305	gi11510977	M. jannaschii predicted coding region M20938 [Methanococcus jannaschii]	74	41	303
4283	1	471	304	gi1520844	orf4 [Bacillus subtilis]	74	58	168
4449	1	3	221	gi1580910	peptide synthetase ORF1 [Bacillus subtilis]	74	54	219
4587	1	458	231	gi11370207	orf6 [Lactobacillus sake]	74	59	228

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4603	1	29	214	gi146208	glutamate synthase large subunit (EC 2.6.1.53) [Escherichia coli] pif[A29617/A29617 glutamate synthase (NADPH) (EC 1.4.1.13) large hain - Escherichia coli]	74	60	186
4670	1	366	184	gi1256135	YbbF [Bacillus subtilis]	74	61	183
5	10	7953	7162	gi143727	putative [Bacillus subtilis]	73	42	792
11	2	2454	1372	gi1466338	dihydroxycarboxylate dehydrogenase [Agrobacterium tumefaciens]	73	55	1083
14	1	2024	1020	gi143373	phosphoribosyl aminimidazole carboxy formyl transferase/inosine monophosphate cyclodiphosphate (Pur-H <sub>2</sub> O) [Bacillus subtilis]	73	54	1005
23	5	5426	4635	gi1468939	meso-2,3-butanediol dehydrogenase (D-acetoin forming) [Klebsiella pneumoniae]	73	58	792
23	17	17379	16360	gi1297060	ornithine cyclodeaminase [Rhizobium meliloti]	73	37	1020
29	2	682	1273	gi1467442	stage V sporulation [Bacillus subtilis]	73	54	582
31	5	6467	4914	gi1414000	ipa-76d gene product [Bacillus subtilis]	73	55	1554
37	8	8658	7402	gi1429259	pept gene product [Bacillus subtilis]	73	59	1257
37	9	7738	7562	gi1468367	alpha-isopropylmalate isomerase (putative) [Rhizomucor ircinellolides]	73	52	177
38	7	3931	4896	gi1405885	yeiN [Escherichia coli]	73	58	966
44	6	5041	4238	gi1240895	unknown [Bacillus subtilis]	73	53	804
44	11	7767	8306	gi142409	mouB gene product [Escherichia coli]	73	50	540
45	3	2439	3080	gi1109685	ProW [Bacillus subtilis]	73	47	642
54	13	14036	13794	gi1413931	ipa-7d gene product [Bacillus subtilis]	73	61	243
59	4	1430	2248	gi147923	threonine dehydratase 2 (EC 4.2.1.16) [Escherichia coli]	73	53	819
65	1	1458	730	gi1677944	AppF [Bacillus subtilis]	73	56	729
80	2	1375	860	gi1460932	murD gene product [Bacillus subtilis]	73	53	516
102	13	10124	11179	gi1580891	3-isopropylmalate dehydrogenase (AA 1 - 365) [Bacillus subtilis] pif[A26522/A26522 3-isopropylmalate dehydrogenase (EC 1.1.1.85) - acillus subtilis]	73	55	1056
109	2	3493	2600	gi1510849	M. jannaschii predicted coding region M0775 [Methanococcus jannaschii]	73	40	894
120	8	4782	5756	gi146970	ribonucleoside triphosphate reductase [Escherichia coli] pif[A47331/A47331 anaerobic ribonucleotide reductase - Escherichia coli]	73	56	975
120	9	5726	6223	gi1204333	anaerobic ribonucleoside-triphosphate reductase [Haemophilus influenzae]	73	62	498

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
132	5	4151	4363	gi 871048	HPSW2 - heavy chain potential motor protein [Giardia intestinalis]	73	43	213
140	6	5952	4324	gi 834107	kdpB [Escherichia coli]	73	59	1629
142	6	7060	5919	gi 410125	ribL gene product [Bacillus subtilis]	73	57	1122
149	4	1866	1717	gi 460892	heparin binding protein-44, MBP-44 (mice, Peptide, 360 aa) p1c [JX0281]JX0281 heparin-binding protein-44 precursor - mouse gi 220434 ORF [Mus musculus] (SUB 2-360)	73	53	150
156	1	1	1431	gi 882504	ORF f560 [Escherichia coli]	73	57	1431
174	6	5352	4525	gi 1146240	ketopantoate hydroxymethyltransferase [Bacillus subtilis]	73	55	828
175	8	5537	5176	gi 854657	Na/H antiporter system ORF3 [Bacillus alcalophilus]	73	56	360
186	5	6593	5493	gi 467477	unknown [Bacillus subtilis]	73	48	1101
249	6	6283	5729	gi 1524397	glycine betaine transporter Opud [Bacillus subtilis]	73	56	555
265	4	1873	2280	gi 39848	U3 [Bacillus subtilis]	73	41	408
270	1	348	582	gi 740461	220 kDa polypeptide (African swine fever virus)	73	53	255
274	4	4283	3618	gi 1204965	hypothetical 23.3 kD protein [Escherichia coli]	73	49	666
279	3	4984	3593	gi 1185288	isochloranate synthase [Bacillus subtilis]	73	58	1392
291	4	1207	1575	gi 1511440	glutamine-fructose-6-phosphate transaminase [Methanococcus jannaschii]	73	63	369
299	2	735	1166	gi 467437	unknown [Bacillus subtilis]	73	58	432
299	5	2050	3234	gi 467439	temperature sensitive cell division [Bacillus subtilis]	73	53	1145
334	1	1237	728	gi 536655	ORF YBR244W [Saccharomyces cerevisiae]	73	43	510
336	2	1827	1036	gi 790943	urea amidolyase [Bacillus subtilis]	73	51	792
374	3	1389	1874	gi 1405451	YneJ [Bacillus subtilis]	73	55	486
433	4	1916	2554	gi 473902	alpha-acetolactate synthase [Lactococcus lactis]	73	54	639
509	2	1795	1028	gi 467483	unknown [Bacillus subtilis]	73	56	768
513	1	1709	918	gi 1146220	[NAD+ dependent glycerol-3-phosphate dehydrogenase [Bacillus subtilis]	73	56	792
533	2	239	733	gi 1510605	hypothetical protein (SP P62297) [Methanococcus jannaschii]	73	44	495
546	2	1148	2815	gi 41748	hcdM protein (AA 1-520) [Escherichia coli]	73	52	1668
549	1	762	382	gi 1314847	GlnA [Bacillus subtilis]	73	57	381
567	1	1346	675	gi 410137	ORFX13 [Bacillus subtilis]	73	58	672

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
716	2	654	1112	gi1256623	exodeoxyribonuclease [Bacillus subtilis]	73	56	459
772	1	3	677	gi142010	Shows 70.3% similarity and 48.6% identity to the EnvM protein of <i>Almonella typhimurium</i> [Aureobae sp.]	73	57	675
774	1	3	209	gi1409286	bmrU [Bacillus subtilis]	73	52	207
782	1	1	402	gi143320	[gap] gene products [Bacillus megaterium]	73	56	402
789	2	451	762	gi11063246	low homology to P14 protein of Hemophilus influenzae and 14.2 kDa protein of <i>Escherichia coli</i> [Bacillus subtilis]	73	56	312
796	1	3	911	gi1453754	ABC transporter [Bacillus subtilis]	73	58	909
806	3	1209	949	gi1433786	cryptophanyl-tRNA synthetase (EC 6.1.1.2) [Bacillus subtilis] pir[370481]wms5 cryptophan-tRNA ligase (EC 6.1.1.2) - <i>Bacillus subtilis</i>	73	51	261
816	2	4839	3097	gi141748	hsmH protein (AA 1-520) [Escherichia coli]	73	52	1743
839	1	798	400	gi1886906	argininosuccinate synthetase [Streptomyces clavuligerus] pir[S57659]S57659 argininosuccinate synthase (EC 6.3.4.5) - <i>Streptomyces clavuligerus</i>	73	59	359
857	1	3	290	gi1348052	acetoin utilization protein [Bacillus subtilis]	73	50	288
1008	1	750	398	gi140100	rodC (tag3) polypeptide (AA 1-746) [Bacillus subtilis] ir[S06049]S06049 rodC protein - <i>Bacillus subtilis</i> p[P3465]TAGF_BACSU TECHNIC ACID BIOSYNTHESIS PROTEIN F.	73	41	393
1018	1	1	213	gi1529357	No definition line found [Caenorhabditis elegans] sp[P46975]STT3_CAEEL OLIGOSACCHARYL TRANSFERASE STT3 SUBUNIT OHOLOG.	73	53	213
1041	1	3	491	gi1142706	comC gene product [Bacillus subtilis]	73	51	489
1174	1	395	204	gi1149513	alpha subunit of laminin 5 [Homo sapiens]	73	60	192
1175	1	655	329	gi1473817	"ORF" [Escherichia coli]	73	57	327
1187	1	3	209	gi1580870	ipa-j/d qoxA gene product [Bacillus subtilis]	73	52	207
1206	1	72	245	gi1144816	formyltetrahydrofolate synthetase (FTHFS) (ttg start codon) (EC 3.4.3) [Mooralla thermacetica]	73	43	174
1454	1	423	241	gi11213253	unknown [Schistosoma japonicum]	73	53	183
1469	1	517	260	gi11303787	YqeG [Bacillus subtilis]	73	55	258
1761	1	374	189	gi19135	Hst26Aa gene product [Drosophila simulans]	73	34	186
1849	1	467	243	gi1162307	DNA topoisomerase II [Trypanosoma cruzi]	73	60	225
2055	1	2	400	gi159381	P47K protein [Rhodococcus erythropolis]	73	34	399
2556	1	2	244	gi1145925	fecB [Escherichia coli]	73	62	243

TABLE 2



S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
2947	2	549	400	gi 1184680	polynucleotide phosphorylase [Bacillus subtilis]	73	51	150
2956	1	746	375	gi 143397	quinol oxidase [Bacillus subtilis]	73	58	372
3037	1	655	329	gi 111091	acetolactate synthase [Bacillus subtilis]	73	55	327
3115	1	385	194	gi 323066	overlapping out-of-phase protein [Eggo plant mosaic virus] sp P20129 V70K_EPMV 70 KD PROTEIN	73	53	192
3603	2	700	527	gi 1419521	glutaryl-CoA dehydrogenase precursor [Mus musculus]	73	48	174
3743	1	798	400	gi 450689	hcdh gene of Escherichia coli product [Escherichia coli] pir S98437 S98437 hcdh protein - Escherichia coli pir S09829 S09629 hypothetical protein A - Escherichia coli [SUB 40-520]	73	54	399
3752	1	640	359	gi 1524193	unknown [Mycobacterium tuberculosis]	73	59	282
3852	1	2	181	gi 216746	D-lactate dehydrogenase [Lactobacillus plantarum]	73	68	180
3914	1	475	239	pir S13490 S134	Hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5) - Chicken (fragment)	73	53	237
3914	2	570	343	gi 524891	unknown [Bacillus subtilis]	73	38	228
4069	1	2	316	gi 40003	oxoglutarate dehydrogenase (NADP) [Bacillus subtilis] p P23129 OD01_BACSU 2-OXOGLOUTARATE DEHYDROGENASE E1 COMPONENT (EC 2.4.2) (ALP1A- KETOGLUTARATE DEHYDROGENASE)	73	55	315
4165	1	715	365	gi 1039521	glutaryl-CoA dehydrogenase precursor [Mus musculus]	73	48	351
4196	1	1	177	gi 409660	deoxyribose-phosphate aldolase [Bacillus subtilis] pir S94955 S94955 deoxyribose-phosphate aldolase (EC 4.1.2.4) - acillus subtilis	73	60	177
4202	1	572	378	gi 524891	unknown [Bacillus subtilis]	73	38	195
4314	1	2	193	gi 436797	N-acyl-L-amino acid amidohydrolase [Bacillus stearothermophilus] sp P37112 AMA_BACST N-ACYL-L-AMINO ACID AMIDOHYDROLASE (EC 5.1.14) (AMINOACYLASE)	73	47	192
4393	1	3	263	gi 216267	ORF2 [Bacillus megaterium]	73	47	261
35	2	903	1973	gi 1146196	phosphoglycerate dehydrogenase [Bacillus subtilis]	72	53	1071
38	22	19094	17877	gi 602031	similar to trimethylamine DH [Mycoplasma capricolum] pir S49950 S49950 probable trimethylamine dehydrogenase (EC 5.99.7) - Mycoplasma capricolum (SGC3) (fragment)	72	54	1218
38	23	18134	19162	gi 413968	ipa-44d gene product [Bacillus subtilis]	72	54	1029
44	19	11895	12953	gi 516272	unknown [Bacillus subtilis]	72	49	1059
48	7	6248	7117	gi 43499	pyruvate synthase [Halobacterium halobium]	72	49	870
50	7	6563	5691	gi 1205399	proton glutamate symport protein [Haemophilus influenzae]	72	53	873

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
53	9	10521	9259	gi1303956	YqjG (Bacillus subtilis)	72	52	1263
56	23	29549	12995	gi1467471	unknown (Bacillus subtilis)	72	47	447
69	4	5298	4123	gi1354725	pfoS/A (Treponema pallidum)	72	46	1176
69	5	4377	4982	gi1904198	hypothetical protein (Bacillus subtilis)	72	43	606
73	1	2	856	gi1142997	glycerol uptake facilitator (Bacillus subtilis)	72	59	855
98	13	9371	10258	gi1467435	unknown (Bacillus subtilis)	72	50	888
127	1	1	1593	gi1217144	alanine carrier protein (thermophilic bacterium PS1) pir[A45111][A45111]	72	56	1593
131	1	5197	2600	gi1153952	alanine transport protein - thermophilic bacterium PG-3	72	53	2598
141	4	1040	1978	gi11405446	polymerase III polymerase subunit (dnaE) (Salmonella typhimurium)	72	53	2598
149	8	2819	2535	gi1606234	pir[A45915][A45915 DNA-directed DNA polymerase (EC 2.7.7.7) III] ipha chain - Salmonella typhimurium	72	54	939
149	17	5472	5245	gi11304472	transketolase (Bacillus subtilis)	72	44	285
154	1	1	210	gi1205620	secY (Escherichia coli)	72	55	228
155	1	2207	1320	gi1391610	DNA polymerase (unidentified phycodnavirus clone OT4)	72	40	210
180	1	2	328	gi1337630	ferritin like protein (Haemophilus influenzae)	72	57	888
184	3	1145	3553	gi11205110	farnesyl diphosphate synthase (Bacillus stearothermophilus)	72	62	327
195	2	1923	1279	gi11001730	pir[3X0257][3X0257 geranyltransferase (EC 2.5.1.10) - Bacillus stearothermophilus	72	49	2409
206	13	14646	15869	gi11064807	virulence associated protein homolog (Haemophilus influenzae)	72	45	665
209	2	462	932	gi11204666	hypothetical protein (Synchocystis sp.)	72	50	1224
215	2	764	522	gi1881513	ORTMININE AMINOTRANSFERASE (Bacillus subtilis)	72	60	471
224	1	2	790	gi1949974	hypothetical protein (GB:X73124_53) (Haemophilus influenzae)	72	63	243
233	1	1526	765	gi11408493	insulin receptor homolog (Drosophila melanogaster) pir[S57245][S57245	72	61	243
240	1	220	1485	gi1537049	insulin receptor homolog - fruit fly (Drosophila melanogaster) (SUB 46-2146)	72	46	1338
245	1	3	1340	gi11204578	lactose repressor (Staphylococcus xylosum)	72	54	789
245	1	3	1340	gi11204578	homologous to SwissProt:VIDA_ECOLI hypothetical protein (Bacillus subtilis)	72	52	762
245	1	3	1340	gi11204578	ORF_0470 (Escherichia coli)	72	52	1266
245	1	3	1340	gi11204578	hypothetical protein (GB:U06949_1) (Haemophilus influenzae)	72	46	1338

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
259	2	2108	1245	gi1340128	ORF1 [Staphylococcus aureus]	72	59	864
304	2	285	1094	gi1205330	glutamine-binding periplasmic protein [Haemophilus influenzae]	72	52	810
307	10	5326	5039	gi1070015	protein-dependent [Bacillus subtilis]	72	53	288
315	1	517	260	gi143399	quinol oxidase [Bacillus subtilis]	72	55	258
316	11	9622	9308	gi1204445	hypothetical protein (SP127857) [Haemophilus influenzae]	72	56	315
337	3	926	1609	gi1487433	citrate synthase II [Bacillus subtilis]	72	55	684
364	7	12538	10493	gi1510643	ferrous iron transport protein B [Methanococcus jannaschii]	72	53	2046
409	2	340	1263	gi1402944	orfRM1 gene product [Bacillus subtilis]	72	49	924
441	3	2177	1590	gi1312379	highly conserved among subacteria [Clostridium acetobutylicum]	72	48	588
453	6	2634	2505	pir180601 BXSA	pir180601 BXSA [antibacterial protein 3 - Staphylococcus haemolyticus]	72	70	150
460	1	2	625	gi1016162	ABC transporter subunit [Cyanophora paradoxa]	72	51	614
463	1	3253	1628	gi1666014	The polymorphism (RFLP) of this gene is associated with susceptibility to essential hypertension. The SA gene product has light homology to acetyl-CoA synthetase [Homo sapiens]	72	60	1626
480	4	1047	3466	gi1433992	ATP synthase subunit epsilon [Bacillus subtilis]	72	53	420
502	1	1086	586	gi1310859	ORF2 [Synecococcus sp.]	72	50	501
519	1	81	1184	gi1303704	YRKE [Bacillus subtilis]	72	54	1104
559	1	3	746	gi1107530	leuO gene product [Campylobacter coli]	72	56	744
575	1	1142	573	gi1303866	VQGS [Bacillus subtilis]	72	56	570
671	1	2	592	gi1204497	protein-export membrane protein [Haemophilus influenzae]	72	44	591
679	2	295	1251	gi1563258	virulence-associated protein 2 [Dichelobacter nodosus]	72	52	957
687	2	295	957	gi1146214	44% identical amino acids with the Escherichia coli sbaB suppressor putative [Bacillus subtilis]	72	49	663
837	1	1	435	gi1146183	putative [Bacillus subtilis]	72	54	435
868	1	150	788	gi1377842	unknown [Bacillus subtilis]	72	55	639
922	1	130	432	gi1088269	unknown protein [Azotobacter vinelandii]	72	58	303
941	1	2	238	gi153929	NAOPH-sulfite reductase flavoprotein component [Salmonella typhimurium]	72	49	237
980	1	840	421	gi1853767	UDP-N-acetylglucosamine 1-carboxyvinyltransferase [Bacillus subtilis]	72	59	420

TABLE 2

5. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
1209	1	383	213	gi1144735	neurotoxin type B [Clostridium botulinum]	72	44	171
1469	2	671	474	gi11205458	hypothetical protein (CB:D26562.47) (Haemophilus influenzae)	72	63	198
1956	1	727	365	gi1154409	hexosephosphate transport protein (Salmonella typhimurium) pfr E41853 E41853 hexose phosphate transport system regulatory protein uhpA - Salmonella typhimurium	72	44	363
2101	1	3	401	gi11303950	YqjY [Bacillus subtilis]	72	50	399
2503	1	569	399	gi1149713	formate dehydrogenase (Methanobacterium formicicum) pfr A42712 A42712 formate dehydrogenase [EC 1.2.1.2] - Methanobacterium formicicum	72	56	171
2967	1	3	155	gi11212729	YqjZ [Bacillus subtilis]	72	46	153
3004	1	367	185	gi11665999	hypothetical protein [Bacillus subtilis]	72	55	183
3109	1	278	141	gi11413968	ipa-44d gene product [Bacillus subtilis]	72	45	138
3171	1	3	287	gi11515938	glutamate synthase (ferredoxin) (Synecocystis sp.) pfr S46957 S46957 glutamate synthase (ferredoxin) [EC 1.4.1.1] - Synecocystis sp.	72	52	285
3771	1	26	367	gi11408501	homologous to N-acetyl-L-amino acid amidohydrolase of Bacillus stearothermophilus [Bacillus subtilis]	72	63	362
3951	1	1	222	gi11500409	M. jannaschii predicted coding region M1519 [Methanococcus jannaschii]	72	38	222
4190	1	721	362	gi1139956	ligG [Bacillus subtilis]	72	57	360
4444	1	3	347	gi11009366	respiratory nitrate reductase [Bacillus subtilis]	72	55	345
6	2	931	1206	gi11537095	ornithine carbamoyltransferase (Escherichia coli)	71	55	270
11	15	11350	10859	gi11532309	25 kDa protein (Escherichia coli)	71	47	492
19	2	1248	2435	gi11244574	[O-alanine:D-alanine ligase (Enterococcus hirae)	71	52	1188
21	2	898	1488	gi1149629	anthranilate synthase component 2 [Leptospira biflexa] pfr C32840 C32840 anthranilate synthase (EC 4.1.3.27) component II Leptospira biflexa	71	45	591
34	1	1	567	gi11303983	YqkF [Bacillus subtilis]	71	59	567
37	3	3192	2806	gi11209681	glutamate-rich protein [Bacillus firmus]	71	50	387
38	18	12250	12462	gi11927645	arginyl endopeptidase (Porphyromonas gingivalis)	71	50	213
39	3	1246	4431	pfr S09411 S094	spoIIIE protein - Bacillus subtilis	71	49	3186
53	14	15770	14760	gi1142613	branched chain alpha-keto acid dehydrogenase E1-alpha [Bacillus subtilis]	71	58	1011
54	11	13461	12625	gi1141014	lont repressor [Bacillus subtilis]	71	46	837

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
57	7	7152	5860	gi 508175	ETC domain of PTS-dependent Cat transport and phosphorylation <i>Escherichia coli</i>	71	48	1293
57	18	11097	14334	gi 1063247	high homology to flavohemoprotein (Haemoglobin-like protein) of <i>Alcaligenes eutrophus</i> and <i>Saccharomyces cerevisiae</i> [ <i>Bacillus subtilis</i> ]	71	56	438
62	16	9831	10955	gi 1303926	YqjG [ <i>Bacillus subtilis</i> ]	71	54	1125
70	12	8505	8966	gi 147198	phtE protein [ <i>Escherichia coli</i> ]	71	38	462
86	5	2394	2089	gi 904205	hypothetical protein [ <i>Bacillus subtilis</i> ]	71	51	306
96	7	7601	8269	gi 705991	hypothetical protein [ <i>Bacillus subtilis</i> ]	71	49	669
100	6	4422	5931	gi 1060848	Opine dehydrogenase [ <i>Arthrobacter</i> sp.]	71	45	1110
101	1	1062	532	gi 1143089	lep protein [ <i>Bacillus subtilis</i> ]	71	41	531
109	18	15312	15695	gi 413985	lipa-61d gene product [ <i>Bacillus subtilis</i> ]	71	57	384
113	1	630	316	gi 663254	probable protein kinase [ <i>Saccharomyces cerevisiae</i> ]	71	57	315
114	5	6598	5603	gi 141156	membrane bound protein [ <i>Bacillus subtilis</i> ]	71	40	996
133	2	3087	1723	gi 1303913	YqhX [ <i>Bacillus subtilis</i> ]	71	53	1365
149	19	6335	5895	gi 529650	G4OP [ <i>Bacteriophage SP21</i> ]	71	51	441
154	5	3635	3087	gi 425488	repressor protein [ <i>Streptococcus sobrinus</i> ]	71	47	549
164	11	11354	11649	gi 49318	ORF4 gene product [ <i>Bacillus subtilis</i> ]	71	52	196
169	5	1936	2745	gi 1403403	unknown [ <i>Mycobacterium tuberculosis</i> ]	71	56	810
193	2	272	1234	gi 1303788	YqjH [ <i>Bacillus subtilis</i> ]	71	49	963
205	1	1743	895	gi 1215694	GlnQ [ <i>Mycoplasma pneumoniae</i> ]	71	46	849
233	4	1849	2022	gi 613732	ORF1 [ <i>Campylobacter jejuni</i> ]	71	50	174
237	7	4501	5169	gi 149384	HistE [ <i>Lactococcus lactis</i> ]	71	54	669
272	4	2848	2273	gi 709993	hypothetical protein [ <i>Bacillus subtilis</i> ]	71	48	576
274	2	618	1496	gi 143035	[NAD(P)H:glutamate]-transfer RNA reductase [ <i>Bacillus subtilis</i> ] p1r[A35252]A35252 5-aminolevulinate synthase (EC 2.3.1.37) - <i>ecillus subtilis</i>	71	53	879
276	5	3349	2720	gi 103562	ORF210 [ <i>Escherichia coli</i> ]	71	50	630
287	1	116	660	gi 110634	20 kDa protein [ <i>Streptococcus gordonii</i> ]	71	53	525
288	6	3322	2771	gi 1256625	putative [ <i>Bacillus subtilis</i> ]	71	47	552

TABLE 2

S. aureus - putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	Accession	Ident	Length (nt)
301	6	3492	2461	gi1467417	similar to lysine decarboxylase [Bacillus subtilis]		71	57
306	4	6607	5222	gi1156618	transport protein [Bacillus subtilis]		71	56
307	2	1536	925	gi1632683	orfC [Mycoplasma capricolum]		71	45
310	5	5793	5146	gi1348052	acetoin utilization protein [Bacillus subtilis]		71	51
322	1	2	1303	gi11001819	hypothetical protein [Synechocystis sp.1]		71	46
333	4	4171	3995	gi1467473	unknown [Bacillus subtilis]		71	57
350	2	548	922	gi1551879	ORF 1 [Lactococcus lactis]		71	55
375	4	1860	3071	gi1467447	unknown [Bacillus subtilis]		71	57
380	5	1560	2102	gi1142557	ATP synthase b subunit [Bacillus megaterium]		71	41
414	2	251	637	gi1580904	homologous to E. coli rnpA [Bacillus subtilis]		71	49
424	1	335	1354	gi1581305	L-lactate dehydrogenase [Lactobacillus plantarum]		71	57
436	4	3701	3270	gi1580501	phosphoribosylanthranilate isomerase (EC 5.3.1.24) - Bacillus subtilis (fragment)		71	66
482	1	1280	1280	gi1410142	ORF18 [Bacillus subtilis]		71	49
525	3	2272	1844	gi1143370	phosphoribosylpyrophosphate amidotransferase (Pur-P; EC 2.4.2.14) Bacillus subtilis		71	56
529	4	2739	2047	gi1606150	ORF1309 [Escherichia coli]		71	41
563	1	22	969	gi11237015	ORF4 [Bacillus subtilis]		71	53
581	1	506	255	gi11301730	T25G3.2 [Caenorhabditis elegans]		71	47
612	2	1068	913	gi1153968	flabrinase 2 [Salmonella typhimurium]		71	55
613	1	1	654	gi1466776	lysine specific permease [Escherichia coli]		71	50
618	1	1243	623	gi11146238	poly(A) polymerase [Bacillus subtilis]		71	52
630	1	1170	586	gi11486243	unknown [Bacillus subtilis]		71	53
691	1	1126	641	gi11289260	comE ORF1 [Bacillus subtilis]		71	51
694	2	149	427	gi112971	NADH dehydrogenase subunit V (AA 1-605) [Gallus gallus] ic1S10197 S10197 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain - chicken mitochondrion (SOL)		71	47
715	2	169	777	gi1103830	YqfL [Bacillus subtilis]		71	53
746	2	1473	970	gi14377843	unknown [Bacillus subtilis]		71	52

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
748	1	1437	802	gi14405459	[YaeS (Bacillus subtilis)]	71	49	636
753	1	1018	524	gi1510389	[M. jannaschii predicted coding region MJ0296 (Methanococcus jannaschii)]	71	53	495
761	1	3	215	gi1475972	[pentafunctional enzyme (Pneumocystis carinii)]	71	47	213
783	1	1203	703	gi1536655	[ORF YBR244w (Saccharomyces cerevisiae)]	71	52	501
900	3	1292	987	gi1204326	[tRNA delta(2)-isopentenylpyrophosphate transferase (Haemophilus influenzae)]	71	48	306
906	1	116	286	gi1419075	[cbiM gene product (Methanobacterium thermoautotrophicum)]	71	50	171
931	1	973	488	gi1893358	[PgsA (Bacillus subtilis)]	71	56	486
1041	1	2	262	gi1408507	[pyrimidine nucleoside transport protein (Bacillus subtilis)]	71	45	261
1070	1	2	172	gi1709993	[hypothetical protein (Bacillus subtilis)]	71	46	171
1176	1	57	365	gi151259	[HMG-CoA reductase (EC 1.1.1.88) (Pseudomonas mavalonii) pfr1A46756/A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.]	71	49	309
1181	1	366	184	gi146971	[lepIP gene product (Staphylococcus epidermidis)]	71	50	183
1281	1	3	290	gi153016	[ORF 419 protein (Staphylococcus aureus)]	71	50	288
1348	1	456	229	gi1402883	[orfC (Mycoplasma capricolum)]	71	48	228
2002	1	756	379	gi11008177	[ORF YJ1046w (Saccharomyces cerevisiae)]	71	48	378
2119	1	2	217	gi11046088	[arginyl-tRNA synthetase (Mycoplasma genitalium)]	71	50	216
2418	1	3	320	gi1499771	[M. jannaschii predicted coding region MJ0936 (Methanococcus jannaschii)]	71	57	318
2561	1	2	187	gi1312443	[carbamoyl-phosphate synthase (glutamine-hydrolysing) (Bacillus aldolyticus)]	71	57	186
2999	2	67	306	gi1710020	[nitrite reductase (nirB) (Bacillus subtilis)]	71	43	240
3033	1	2	184	gi1262335	[YmaA (Bacillus subtilis)]	71	57	183
3584	1	3	338	gi1401716	[beta-isopropylmalate dehydrogenase (Neurospora crassa)]	71	55	336
3715	2	743	399	gi1563952	[glucanase permease (Bacillus licheniformis)]	71	59	345
3785	1	770	387	gi147382	[acyl-CoA-dehydrogenase (Streptomyces purpurascens)]	71	57	384
3875	1	541	272	gi11001541	[hypothetical protein (Synechocystis sp.)]	71	38	270
4135	1	637	320	gi1342695	[S-adenosyl-L-methionine:uroporphyrinogen III methyltransferase (Bacillus megaterium)]	71	52	318
4249	1	63	239	gi1205363	[deoxyribose aldolase (Haemophilus influenzae)]	71	63	177
4508	1	530	267	gi11191667	[vitellogenin (Anolis pulchellus)]	71	46	264

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
6	3	1237	2721	gi1321788	arginine ornithine antiporter [Clostridium perfringens]	70	54	1485
11	11	6572	7486	gi1216854	p47K [Pseudomonas chlororaphis]	70	41	915
12	1	2890	1481	gi1467330	replicative DNA helicase [Bacillus subtilis]	70	49	1410
15	1	1756	893	gi1451216	Mannosephosphate isomerase [Streptococcus mutans]	70	46	864
15	2	1277	1050	gi1476092	unknown [Bacillus subtilis]	70	50	228
17	2	2112	1350	gi1145402	choline dehydrogenase [Zscherichia coli]	70	52	783
21	1	2	925	gi1149516	anthranilate synthase alpha subunit [Lactococcus lactis] p1r.S15124	70	50	924
					anthranilate synthase (EC 4.1.3.27) alpha chain - actococcus lactis subsp. lactis			
25	7	5580	6251	gi11389549	ORF3 [Bacillus subtilis]	70	52	672
31	6	6071	7423	gi11301875	Yqha [Bacillus subtilis]	70	51	1353
36	2	959	1594	gi1500755	laechyl purine glycosylase [Mus musculus]	70	47	616
38	6	4901	5860	gi11408507	pyrimidine nucleoside transport protein [Bacillus subtilis]	70	44	960
44	8	5312	5989	gi11006620	hypothetical protein [Synchocystis sp.]	70	49	678
46	10	8950	10020	gi11403126	czcd gene product [Alcaligenes eutrophus]	70	45	1071
52	2	2727	1900	gi11486247	unknown [Bacillus subtilis]	70	53	828
52	6	4048	4656	gi1244501	esterase II:carboxylesterase (EC 3.1.1.1) [Pseudomonas fluorescens, eptido, 218 aa]	70	50	609
56	8	8460	9962	gi11339951	small subunit of NADH-dependent glutamate synthase [Plectonama boryanum]	70	51	1503
62	1	48	290	gi1142702	A competence protein 2 [Bacillus subtilis]	70	47	243
64	1	1080	541	gi11204377	hemolysin biosynthesis protein [Haemophilus influenzae]	70	47	540
70	5	5139	3595	gi11204834	2',3'-cyclic-nucleotide 2'-phosphodiesterase [Haemophilus influenzae]	70	47	1545
91	4	7793	5466	gi11886471	methionine synthase [Catharanthus roseus]	70	56	2328
96	5	8754	7255	gi11390961	alkaline phosphatase (EC 3.1.3.1) III precursor - Bacillus subtilis	70	54	1500
110	2	767	1300	gi1145294	adenine phosphoribosyl-transferase [Escherichia coli]	70	51	534
116	6	7026	7976	gi1143607	sporulation protein [Bacillus subtilis]	70	50	951
121	8	6401	6988	gi1107528	ttg start [Campylobacter coli]	70	45	588
131	8	6842	7936	gi11150454	prolidase PpQ [Lactobacillus delbrueckii]	70	48	1095

TABLE 2



S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match nCession	Match gene name	A sim	A idont	length (nt)
135	1	2	1489	gi1311109	putative membrane-bound protein with four times repetition of ro-Ser-Ala at the N-terminus; function unknown [Alcaligenes utrophus]	70	49	1488
138	3	418	714	gi1504181	hypothetical protein [Bacillus subtilis]	70	46	297
164	8	9344	9874	gi149315	ORF1 gene product [Bacillus subtilis]	70	47	531
164	16	15626	16618	gi1205212	hypothetical protein (GB:D1048)_18 [Haemophilus influenzae]	70	50	993
205	2	2735	1803	gi1215695	peptide transport system protein SapF homolog; SapF homolog [Mycoplasma pneumoniae]	70	47	933
209	3	910	1386	gi11204655	hypothetical protein (GB:X73124_26) [Haemophilus influenzae]	70	48	477
246	3	340	756	gi1215098	excisionase [Bacteriophage 154a]	70	46	417
263	7	7876	6749	gi1142540	aspartokinase II [Bacillus sp.]	70	51	1128
268	3	3212	4117	gi1140128	ORF1 [Staphylococcus aureus]	70	50	906
302	6	3201	3827	gi1147782	ruvA protein (atg start) [Escherichia coli]	70	46	627
302	10	5879	7051	pir[C18530]C185	guanine tRNA-ribosyltransferase (EC 2.4.2.29) - Escherichia coli	70	55	1173
313	1	2520	1416	gi11205934	aminopeptidase a/1 [Haemophilus influenzae]	70	46	1107
335	2	379	669	gi11070013	protein-dependent [Bacillus subtilis]	70	48	291
403	1	1255	629	gi1733147	GuaF [Xanthomonas campestris]	70	33	627
414	10	4770	9273	gi11204752	high affinity ribose transport protein [Haemophilus influenzae]	70	52	503
449	1	2	1243	gi11619724	HgtE [Bacillus firmus]	70	43	1242
472	1	637	320	gi1727145	open reading frame; putative [Bacillus amyloquelactone] pir[B29091]B29091 - Bacillus myoliquefaciens (fragment)	70	41	318
480	2	727	1608	gi1142560	ATP synthase gamma subunit [Bacillus megaterium]	70	44	882
524	1	2	307	gi1602292	ROH2 protein [Brassica napus]	70	45	306
525	1	823	413	gi1143372	phosphoribosyl glycinate formyltransferase (Pur-N) [Bacillus subtilis]	70	52	411
565	4	1625	2552	gi1881434	ORP [Bacillus subtilis]	70	51	1074
607	4	829	1284	gi1511524	hypothetical protein (SP:P17002) [Methanococcus jannaschii]	70	50	456
633	1	1383	703	gi1431231	luciferase [Bacillus caldolyticus]	70	53	681
646	3	1683	1309	gi1467340	unknown [Bacillus subtilis]	70	49	375
663	1	830	417	gi11303873	Y11Z [Bacillus subtilis]	70	40	414

TABLE 2

1 aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
641	1	1488	781	gi 1001678	hypothetical protein [Synecocystis sp.]	70	53	768
708	1	2	448	sp P33940 YOH_	HYPOTHETICAL 56.3 KD PROTEIN IN ECO-ALKB INTERGENIC REGION	70	51	447
725	1	51	722	gi 1001644	hypothetical protein [Synecocystis sp.]	70	48	632
776	1	1371	767	gi 145165	putative [Escherichia coli]	70	47	585
834	1	250	783	gi 552971	NADH dehydrogenase (ndhF) [Vicia faba]	70	47	534
865	2	1585	1379	gi 1204636	ATP-dependent helicase [Haemophilus influenzae]	70	45	207
894	1	535	269	gi 467364	DNA binding protein (probae) [Bacillus subtilis]	70	41	267
919	1	3	317	gi 114847	[ClnA [Bacillus subtilis]	70	40	315
944	1	3	572	gi 709991	hypothetical protein [Bacillus subtilis]	70	44	570
988	2	772	605	gi 142441	ORF 3; putative [Bacillus subtilis]	70	50	168
1055	1	3	335	gi 529755	apeC [Streptococcus pyogenes]	70	37	333
1093	1	2	904	gi 853754	ABC transporter [Bacillus subtilis]	70	49	903
1109	1	2	310	gi 1001827	hypothetical protein [Synecocystis sp.]	70	42	309
1220	1	468	235	sp S23416 S234	epib protein - Staphylococcus epidermidis	70	40	234
1279	1	73	348	gi 153015	PemA protein [Staphylococcus aureus]	70	47	276
1316	1	195	542	sp P31776 P8PA_	PENICILLIN-BINDING PROTEIN 1A (PBP-1A) [PENICILLIN-BINDING PROTEIN A]	70	50	348
1537	2	232	402	gi 1146181	putative [Bacillus subtilis]	70	50	171
1574	1	451	272	gi 219630	endothelin-A receptor [Homo sapiens]	70	47	180
1640	1	690	346	gi 1146243	22.4% identity with Escherichia coli DNA-damage inducible protein ...; putative [Bacillus subtilis]	70	46	345
2504	1	2	286	gi 495179	transmembrane protein [Lactococcus lactis]	70	51	285
3061	1	564	301	gi 508175	ELIC domain of PTS-dependent Gal transport and phosphorylation Escherichia coli	70	44	264
3128	1	2	199	gi 1340096	unknown [Mycobacterium tuberculosis]	70	51	198
3218	1	3	488	gi 151938	glutamate synthase (ferredoxin) [Synecocystis sp.] pir S46957 S46957	70	50	480
3323	1	794	399	gi 1154891	ATP binding protein [Phormidium laminosum]	70	52	396
3679	1	599	399	gi 529385	chromosome condensation protein [Caenorhabditis elegans]	70	30	201

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
3841	1	706	gi1208965	hypothetical 23.3 kd protein [Escherichia coli]	70	47	309
3929	1	3	gi149435	purative [Lactococcus lactis]	70	49	399
4044	1	595	gi1602031	similar to trimethylamine DH [Mycoplasma capricolum] pif154950[S49950 probable trimethylamine dehydrogenase [EC 5.99.7] - Mycoplasma capricolum (SGC3)] (fragment)	70	40	222
4329	1	558	gi1339951	small subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	70	49	279
4422	1	576	gi1296464	ATPase [Lactococcus lactis]	70	57	288
4647	1	361	gi166412	NADH-glutamate synthase [Medicago sativa]	70	59	162
16	8	7571	gi1499620	M. jannaschii predicted coding region MJ0798 [Methanococcus jannaschii]	69	44	1463
16	9	9080	gi1353197	thioredoxin reductase [Escherichia coli]	69	54	954
30	1	1452	gi1204910	hypothetical protein (GB:U14003_302) [Haemophilus influenzae]	69	52	726
38	4	1023	gi1407773	deva gene product [Anabaena sp.]	69	41	276
44	9	5987	gi1205920	molybdate uptake system hydrophilic membrane-bound protein [Haemophilus influenzae]	69	45	609
62	15	9104	gi1385178	unknown [Bacillus subtilis]	69	44	372
66	4	2402	gi1303893	YqjL [Bacillus subtilis]	69	51	402
67	15	14124	gi149647	ORF2 [Listeria monocytogenes]	69	37	494
67	17	14053	gi1305002	ORF135 [Escherichia coli]	69	49	330
67	19	15130	gi1109684	ProV [Bacillus subtilis]	69	45	678
78	3	1447	gi1256633	putative [Bacillus subtilis]	69	53	678
78	4	4513	gi1103398	YqjG [Bacillus subtilis]	69	52	789
85	4	4521	gi129326[E293]	hypothetical protein (pur operon) - Bacillus subtilis	69	32	309
86	6	3253	gi1973332	OrfC [Bacillus subtilis]	69	50	600
95	1	96	gi1766468	4411 antigen, sperm tail membrane antigen-putative sucrose-specific phosphotransferase enzyme II homolog [rice, teosin, Peptide Partial, 72 aa]	69	43	615
100	7	6023	gi1205355	Na+/H+ antiporter [Haemophilus influenzae]	69	39	1404
102	2	2678	gi1561690	isialoglycoprotease [Pasteurella haemolytica]	69	47	1029
103	8	12241	gi1009366	Respiratory nitrate reductase [Bacillus subtilis]	69	54	3705
103	11	14987	gi1710020	nitrite reductase (nirB) [Bacillus subtilis]	69	51	2436

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
112	11	8708	10168	gi 154111	hexosephosphate transport protein [Salmonella typhimurium] p1r D41853 D41853 hexose phosphate transport system protein uhpT - salmonella typhimurium	69	51	1461
112	16	16644	17414	gi 1204435	pyruvate formate-lyase activating enzyme [Haemophilus influenzae]	69	50	771
113	2	33	953	gi 1290509	lo307 [Escherichia coli]	69	43	921
114	2	1537	1058	p1r A42771 A427	reticulocyte-binding protein 1 - Plasmodium vivax	69	39	480
121	6	4309	5310	gi 1154633	NrdF [Bacillus subtilis]	69	53	1002
125	2	267	854	gi 113931	lpa-7d gene product [Bacillus subtilis]	69	43	588
149	27	10666	10400	p1r S28089 S280	hypothetical protein A - yeast [Zygosaccharomyces bisporus] plasmid pSu]	69	39	267
161	1	1598	813	gi 1205538	hypothetical protein (CB:U14003_302) [Haemophilus influenzae]	69	47	786
165	4	2222	4633	gi 140054	phenylalanyl-tRNA synthetase beta subunit (AA 1-804) [Bacillus thilis]	69	52	2412
169	3	1210	1761	gi 1296031	elongation factor Ts [Spirulina platensis]	69	45	552
175	12	8686	8339	gi 732682	Pme protein [Escherichia coli]	69	69	348
190	2	484	1671	sp P17731 H158	HISTIDINOL-PHOSPHATE AMINOTRANSFERASE (EC 2.6.1.9) (IMIDAZOLE ACETOL-PHOSPHATE TRANSAMINASE)	69	48	1188
206	1	5551	2777	gi 41750	hadr protein (AA 1-1033) [Escherichia coli]	69	49	2775
206	4	6038	5796	gi 1256135	YnfF [Bacillus subtilis]	69	48	243
249	1	636	319	gi 1404456	YnfP [Bacillus subtilis]	69	50	316
302	8	4820	5776	gi 1001768	hypothetical protein [Synechocystis sp.]	69	48	957
324	2	7384	3893	gi 1256798	pyruvate carboxylase [Rhizobium etli]	69	53	1492
351	3	2098	1808	gi 1491664	T04H1.4 [Caenorhabditis elegans]	69	30	291
369	3	2075	2105	gi 336458	ORF [Balanoptera acutorostrata]	69	61	231
392	3	1999	2424	gi 556015	ORF [Bacillus subtilis]	69	45	426
410	1	87	779	gi 155611	phosphoglyceromutase [Zymomonas mobilis]	69	58	693
421	1	2085	1129	gi 1276985	arginase [Bacillus caldovelox]	69	54	957
444	8	6713	7741	gi 1221782	purine synthesis repressor [Haemophilus influenzae]	69	40	1029
453	1	828	415	gi 1122758	unknown [Bacillus subtilis]	69	57	414
469	2	3286	2246	gi 1458228	muty homolog [Homo sapiens]	69	44	1041

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	Align	Ident	Length (nt)
509	3	1730	1371	gi 49224	URF 4 (Synecoccus sp.)	69	39	360
520	5	3023	2823	gi 726427	similar to D. melanogaster MST101-2 protein (P.R.SJ4134) Caenorhabditis elegans	69	39	201
531	1	26	760	gi 509672	repressor protein (bacteriophage Tuc2009)	69	33	735
589	1	107	253	gi 169101	17.9 kDa heat shock protein (hsp17.9) (Pisum sativum)	69	52	147
594	2	597	1391	gi 542783	DNA photolyase (Bacillus firmus)	69	48	795
604	4	2476	2114	gi 413930	ipa-6d gene product (Bacillus subtilis)	69	45	361
607	1	2	313	gi 12316103	W08D2.3 (Caenorhabditis elegans)	69	47	312
607	2	590	312	gi 536715	ORF Y8R275c (Saccharomyces cerevisiae)	69	39	279
734	1	864	433	gi 467327	unknown (Bacillus subtilis)	69	44	432
759	1	3	338	gi 1009367	Respiratory nitrate reductase (Bacillus subtilis)	69	50	336
761	2	392	586	gi 3508	Leucyl-tRNA synthetase (cytoplasmic) (Saccharomyces cerevisiae)	69	46	195
802	1	72	1013	gi 143044	ORF YEL160u (Saccharomyces cerevisiae)	69	55	942
816	1	2573	1360	gi 1510268	restriction modification system S subunit (Methanococcus jannaschii)	69	45	1206
838	2	133	387	gi 1255371	coded for by C. elegans cDNA YK349.5; coded for by C. elegans cDNA YK349.3; Similar to guanilate kinase (Caenorhabditis elegans)	69	46	255
851	2	745	1005	gi 248998	secA gene product (Antilimonium sp.)	69	39	261
867	1	535	269	gi 1070014	protein-dependent (Bacillus subtilis)	69	47	267
995	1	954	478	gi 205569	transcription elongation factor (Haemophilus influenzae)	69	53	477
999	1	1009	506	gi 859254	predicted trichorax protein (Drosophila virilis)	69	21	504
1127	1	1315	659	gi 1205434	H. influenzae predicted coding region H1191 (Haemophilus influenzae)	69	56	657
1138	1	248	460	gi 1510646	H. jannaschii predicted coding region MJ0568 (Methanococcus jannaschii)	69	48	213
2928	1	3	401	gi 20503	glutamate permease (Escherichia coli)	69	41	399
3090	1	444	223	gi 1204987	DNA polymerase III, alpha chain (Haemophilus influenzae)	69	36	222
3817	1	2	400	gi 1483199	peptide synthetase (Ancyloctopais mediterranei)	69	45	399
3833	1	667	335	gi 1574193	unknown (Mycobacterium tuberculosis)	69	46	333

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
4079	1	747	400	gi1546918	orfV 3' of comK [Bacillus subtilis, E26, Peptide Partial, 140 aa] pir[S41612]S41612 hypothetical protein Y - Bacillus subtilis ap140198 YHXO_BACSU HYPOTHETICAL PROTEIN IN COMK 3'REGION (ORFY FRAGMENT)	69	64	148
4115	2	215	400	gi1517205	67 kDa Myosin crossreactive streptococcal antigen [Streptococcus pyogenes]	69	59	186
4139	1	1	333	gi1208451	hypothetical protein [Synchocystis sp.]	69	36	333
4258	1	457	230	gi1496158	restriction-modification enzyme subunit M1 [Mycoplasma pulmonis] pir[S49395]S49395 [adM1 protein - Mycoplasma pulmonis (SQC)]	69	41	228
4317	1	90	374	gi1413967	lpa-43d gene product [Bacillus subtilis]	69	44	285
4465	1	1	293	gi1396296	similar to phosphotransferase system enzyme II [Escherichia coli] ap121672 PTMC_ECOLI PTS SYSTEM, FRUCTOSE-LIKE-2 IIC COMPONENT PHOSPHOTRANSFERASE ENZYME II, C COMPONENT	69	49	291
3	1	2102	1193	gi1109685	ProW [Bacillus subtilis]	68	46	1110
15	4	2592	2074	gi1807973	unknown [Saccharomyces cerevisiae]	68	45	519
31	8	6328	872	gi1290642	ATPase [Enterococcus hirae]	68	48	2445
40	2	1115	750	gi1606342	ORF_0622, reading frame open far upstream of start; possible ranseshift, linking to previous ORF [Escherichia coli]	68	55	166
46	9	6886	8415	gi1155276	aldehyde dehydrogenase [Vibrio cholerae]	68	44	1530
48	3	3643	3404	gi1285608	241k polyprotein [Apple stem grooving virus]	68	47	240
48	4	3536	4132	gi1104537	M. genitalium predicted coding region MQ246 [Mycoplasma genitalium]	68	39	597
53	10	11671	10685	gi11301952	YqjA [Bacillus subtilis]	68	46	987
70	9	7346	8155	gi147198	phnE protein [Escherichia coli]	68	40	810
89	4	1899	2966	gi1145173	35 kDa protein [Escherichia coli]	68	43	1068
108	1	2187	1150	gi138722	precursor (aa -20 to 381) [Acinetobacter calcoaceticus] ir[A29277] aldose 1-epimerase (EC 5.1.3.3) - Acinetobacter calcoaceticus	68	57	1038
112	5	2866	1622	gi1153724	MalC [Streptococcus pneumoniae]	68	55	957
116	7	7865	8638	gi1141608	agorulation protein [Bacillus subtilis]	68	48	774
118	3	2484	3698	gi11301805	YqeR [Bacillus subtilis]	68	46	1215
120	2	1424	1594	ap138038 CYSD	SULFITE REDUCTASE (NADPH) FLAVOPROTEIN ALPHA COMPONENT (EC 1.8.1.2) (SIR FP)	68	45	171
129	1	1	1011	gi1396307	argininosuccinate lyase [Escherichia coli]	68	50	1011

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
132	3	1867	gi 516267	ORF2 [Bacillus megaterium]	68	48	873
134	2	848	gi 140545	DNA recombinase [Escherichia coli]	68	50	165
141	2	372	gi 872116	act (stress inducible protein) [Glycine max]	68	36	243
149	7	2454	gi 145774	hsp70 protein (dnaK gene) [Escherichia coli]	68	48	195
155	2	1776	gi 216583	ORF1 [Escherichia coli]	68	36	243
158	3	1826	gi P33940 YQJH	HYPOTHETICAL 34.3 KD PROTEIN IN ECO-ALKB INTERGENIC REGION	68	51	1464
169	6	2749	gi 1403402	unknown [Mycobacterium tuberculosis]	68	46	570
175	10	9158	gi 1072395	phaA gene product [Rhizobium meliloti]	68	51	1794
188	7	4184	gi 1173843	3-ketoacyl-ACP synthase II [Vibrio Harveyi]	68	48	1251
189	3	907	gi 467383	DNA binding protein (probable) [Bacillus subtilis]	68	55	759
206	5	7883	gi 1256138	YbbI [Bacillus subtilis]	68	48	975
206	8	110425	gi 452687	pyruvate decarboxylase [Saccharomyces cerevisiae]	68	48	1752
212	8	3421	gi 1369941	lcl gene product [Bacteriophage B1]	68	39	228
214	8	5457	gi 1420467	ORF YOR196c [Saccharomyces cerevisiae]	68	45	1026
237	4	2507	gi 149381	HlsH [Lactococcus lactis]	68	46	582
241	5	5540	gi 1235684	mevalonate pyrophosphate decarboxylase [Saccharomyces cerevisiae]	68	47	999
262	1	3	gi 150974	4-oxalocrotonate tautomerase [Pseudomonas putida]	68	42	162
262	2	1984	gi 1147744	PSR [Enterococcus hirae]	68	49	867
276	6	3702	gi P30750 ABC_E	ATP-BINDING PROTEIN ABC (FRAGMENT)	68	50	564
306	6	6345	gi 1256617	adenine phosphoribosyltransferase [Bacillus subtilis]	68	53	621
333	3	4599	gi 467473	unknown [Bacillus subtilis]	68	45	750
365	6	5017	gi 1130643	722B3.3 [Caenorhabditis elegans]	68	45	180
376	2	549	gi 1277026	DAPA aminotransferase [Bacillus subtilis]	68	51	1098
405	1	1741	gi 1303917	YgiB [Bacillus subtilis]	68	47	870
406	2	853	gi 1511513	ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii]	68	44	315
426	6	3558	gi 624632	GltB [Escherichia coli]	68	48	168
438	3	108	gi 146923	nitrogenase reductase [Escherichia coli]	68	43	222

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
443	1	476	240	gi 535810	hippuricase [Campylobacter jejuni]	68	42	237
443	2	518	1015	gi 1204742	H. influenzae predicted coding region H10491 [Haemophilus influenzae]	68	48	498
443	5	4447	3779	gi 504660	deoxyribose-phosphate aldolase [Bacillus subtilis] pir S49455 S49455	68	55	669
476	2	240	1184	gi 571345	deoxyribose-phosphate aldolase [E. coli] acillus subtilis	68	45	945
486	2	1876	1046	gi 147328	unknown, similar to E. coli cardiolipin synthase [Bacillus subtilis] sp P45860 YME_BACSU HYPOTHETICAL 58.2 PROTEIN IN NARI-ACDA INTERGENIC REGION	68	45	831
517	3	1764	2084	gi 523809	transport protein [Escherichia coli]	68	64	321
517	3	1764	2084	gi 523809	orf2 [Bacteriophage A2]	68	47	510
517	3	1764	2084	gi 523809	HYPOTHETICAL 6.8 KD PROTEIN IN NRDC-TK INTERGENIC REGION	68	47	510
646	1	914	459	gi 413982	ipa-58r gene product [Bacillus subtilis]	68	52	456
659	3	1668	1901	gi 1407541	C3309.8 [Caenorhabditis elegans]	68	36	234
864	5	1510	1716	gi 145774	hap70 protein (dnaK gene) [Escherichia coli]	68	48	207
920	1	860	432	gi 110416	hypothetical protein (SP_P31466) [Methanococcus jannaschii]	68	54	439
952	1	1096	611	gi C33456	reductase [Leishmania major]	68	46	486
970	1	91	402	gi 1354775	pfoS/R [Treponema pallidum]	68	46	312
1078	1	1064	534	gi C10217	diaminopimelate decarboxylase [Bacillus subtilis]	68	47	531
1079	1	428	216	gi 1335714	Plasmodium falciparum mRNA for aspartaginase-like antigen (clone LTC1) [Plasmodium falciparum]	68	31	213
1058	1	692	348	gi C81649	lepC gene product [Staphylococcus epidermidis]	68	46	345
1096	2	665	465	gi 143434	Rho Factor [Bacillus subtilis]	68	43	201
1308	1	2	694	gi 169939	Group B oligopeptidase PepB [Streptococcus agalactiae]	68	50	693
1679	1	2	238	gi 17205	167 kDa Myosin-crossreactive streptococcal antigen [Streptococcus pyogenes]	68	53	237
2039	1	3	383	gi 153898	transport protein [Salmonella typhimurium]	68	51	381
2077	1	3	326	pir C33496 C334	hisc homolog - Bacillus subtilis	68	47	324
2112	1	613	374	gi 64884	lamin L11 [Xenopus laevis]	68	50	240
2273	1	793	398	gi 581648	lepB gene product [Staphylococcus epidermidis]	68	45	396
2948	1	2	385	gi 216869	branched-chain amino acid transport carrier [Pseudomonas aeruginosa] pir A38534 A38534 branched-chain amino acid transport protein braz	68	41	386
					Pseudomonas aeruginosa			

TABLE 2



S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
2955	1	768	400	gi1904179	hypothetical protein [Bacillus subtilis]	68	49	369
2981	1	572	288	gi1508979	GTP-binding protein [Bacillus subtilis]	68	48	285
3014	1	584	294	gi1524394	ORF-2 upstream of gbaB operon [Bacillus subtilis]	68	45	291
3082	1	336	169	gi1204696	fructose-permease IIC component [Haemophilus influenzae]	68	53	168
3108	1	103	258	gi1217855	heat-shock protein [Arabidopsis thaliana]	68	48	156
3639	1	919	461	gi11510490	nitrate transport permease protein [Methanococcus jannaschii]	68	47	459
3657	1	1	330	gi1155369	PTS enzyme-II fructose [Xanthomonas campestris]	68	48	330
3823	1	780	391	gi1603768	HutI protein, imidazole-5-propionate hydrolase [Bacillus subtilis]	68	54	390
3982	1	2	277	gi1149435	putative [Lactococcus lactis]	68	47	276
4051	1	1	342	gi1450688	hadM gene of Ecoprr1 gene product [Escherichia coli] pir[S38437/S38437 hadM protein - Escherichia coli] pir[S09629/S09629 hypothetical protein A - Escherichia coli] (SUB 40-520)	68	48	342
4089	1	12	209	gi11353678	heavy-metal transporting P-type ATPase [Proteus mirabilis]	68	47	198
4143	1	47	187	gi1603769	HutU protein, urecanase [Bacillus subtilis]	68	55	141
4148	1	2	352	gi1450688	hadM gene of Ecoprr1 gene product [Escherichia coli] pir[S38437/S38437 hadM protein - Escherichia coli] pir[S09629/S09629 hypothetical protein A - Escherichia coli] (SUB 40-520)	68	51	351
4173	1	2	382	gi1041097	Pyruvate Kinase [Bacillus psychrophilus]	68	48	381
4182	1	498	250	gi1411966	ipe-44d gene product [Bacillus subtilis]	68	50	249
4362	2	148	318	gi1450688	hadM gene of Ecoprr1 gene product [Escherichia coli] pir[S38437/S38437 hadM protein - Escherichia coli] pir[S09629/S09629 hypothetical protein A - Escherichia coli] (SUB 40-520)	68	44	171
5	11	9493	8300	gi1143727	putative [Bacillus subtilis]	67	46	1194
31	11	10318	9833	gi1216746	D-lactate dehydrogenase [Lactobacillus plantarum]	67	41	486
32	1	1560	1155	gi1098557	renal sodium/dicarboxylate cotransporter (Homo sapiens)	67	46	1596
32	5	4945	4145	gi1510720	prephenate dehydratase [Methanococcus jannaschii]	67	51	801
36	5	5350	4268	gi146216	45% identity with the product of the ORF6 gene from the Erwinia herbicola carotenoid biosynthesis cluster; putative [Bacillus subtilis]	67	58	1083
44	7	4492	5304	gi11006621	hypothetical protein [Synchocystis sp.]	67	43	813

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
56	7	3943	8481	gi 394131	glutamate synthase large subunit precursor (Acropirillum brasiliense) [gi 346602 346602 glutamate synthase (NADPH) (EC 1.4.1.13) alpha chain - Acropirillum brasiliense]	67	52	4539
56	12	11923	14678	gi 1000453	TrpR (Bacillus subtilis)	67	48	756
62	8	5092	4757	gi 1133949	orf3 (Bacillus C-125, alkali-sensitive mutant 18224, Peptide Mutant, 112 aa)	67	45	116
62	10	7570	6338	gi 154655	Na/H antiporter system (Bacillus alcalophilus)	67	49	1233
99	3	2119	3321	gi 1208349	hypothetical protein (GB:090212_3) (Haemophilus influenzae)	67	50	1201
102	9	5695	7176	gi 149432	putative (Lactococcus lactis)	67	51	1482
103	13	14549	14049	gi 3408497	LPD gene product (Bacillus subtilis)	67	48	501
109	15	14821	13982	gi 413976	ipe-52r gene product (Bacillus subtilis)	67	49	840
109	17	14811	15194	gi 413983	ipe-59d gene product (Bacillus subtilis)	67	29	384
121	4	1713	2153	gi 1262335	YmaA (Bacillus subtilis)	67	54	441
122	1	1	1149	gi 143047	ORF8 (Bacillus subtilis)	67	35	1149
124	5	4060	3518	gi 556885	Unknown (Bacillus subtilis)	67	47	543
131	2	4384	3589	gi 1046081	hypothetical protein (GB:D3845_10) (Mycoplasma genitalium)	67	30	996
140	3	2899	2297	gi 146545	kdpC (Escherichia coli)	67	45	603
142	4	5109	4198	gi 1212775	GTP cyclohydrolase II (Bacillus amyloliquefaciens)	67	55	1212
147	5	2913	2374	gi 1303709	YrkJ (Bacillus subtilis)	67	44	540
152	8	6341	6673	gi 1377841	Unknown (Bacillus subtilis)	67	48	333
161	4	2720	3763	gi 496319	SphX (Synecococcus sp.)	67	47	1044
163	6	1989	3428	gi 595681	2-oxoglutarate/malate translocator (Spinacia oleracea)	67	47	1440
193	3	1351	1626	gi 1511101	shikimate 5-dehydrogenase (Methanococcus jannaschii)	67	53	276
200	2	917	2179	gi 142439	ATP-dependent nuclease (Bacillus subtilis)	67	48	1263
206	10	12445	112801	sp P37347 YECD_	HYPOTHETICAL 21.8 KD PROTEIN IN ASPS 5' REGION	67	47	357
206	11	13047	14632	gi 732813	branched-chain amino acid carrier (Lactobacillus delbrueckii)	67	46	1386
208	2	1321	809	gi 1033037	100 kDa heat shock protein (Hsp100) (Leishmania major)	67	36	513
238	3	1039	2052	gi 809542	CbrB protein (Erwinia chrysanthemi)	67	42	1014

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
246	2	176	367	gi 215098	exclusionase [Bacteriophage 154a]	67	37	192
276	2	2260	1412	gi 303560	ORF271 [Escherichia coli]	67	50	849
297	6	2223	3056	gi 142784	CtaA protein [Bacillus firmus]	67	46	834
307	7	5220	4186	gi 1070013	protein-dependent [Bacillus subtilis]	67	43	1035
316	1	36	1028	gi 1161061	dioxigenase [Methylobacterium extorquens]	67	52	993
324	3	5650	5030	gi 1469784	putative cell division protein ftsW [Enterococcus hirae]	67	49	621
336	1	524	264	gi 173122	urea amidolyase [Saccharomyces cerevisiae]	67	45	261
360	1	1008	1394	gi 109053	11S10YU-TNNA SYNTHETASE [EC 6.1.1.21] (HISTIDINE--TNNA LIGASE) (HISMG)	67	47	1287
364	3	4890	3592	gi 151259	HMG-CoA reductase [EC 1.1.1.88] (Pseudomonas mavaltonii) pir[A44756]A44756 hydroxymethylglucaryl-CoA reductase [EC 1.1.1.88] Pseudomonas sp.	67	46	1299
365	3	2940	2113	gi 1296823	orf2 gene product [Lactobacillus helveticus]	67	47	828
367	2	325	918	gi 1039479	ORFU [Lactococcus lactis]	67	47	594
395	3	666	1271	gi 1204516	hypothetical protein (GB:U00016.4) [Haemophilus influenzae]	67	55	606
415	1	1800	901	gi 182579	CG Site No. 29739 [Escherichia coli]	67	46	900
419	1	1799	803	gi 520752	putative [Bacillus subtilis]	67	48	897
474	1	2	796	gi 146906	argininosuccinate synthetase [Streptomyces clavuligerus] pir[S57659]S57659 argininosuccinate synthase [EC 6.3.4.5] - treptomyces clavuligerus	67	49	775
485	2	1921	2226	gi 143434	Rho factor [Bacillus subtilis]	67	43	306
596	1	1728	865	gi 1030853	Xqpf [Bacillus subtilis]	67	47	864
700	1	433	218	gi 1204628	hypothetical protein (SP:P21498) [Haemophilus influenzae]	67	47	216
806	2	249	647	gi 677917	AppC [Bacillus subtilis]	67	51	399
828	2	340	900	gi 777761	IrrA [Synecoccus sp.]	67	37	561
833	1	1407	916	gi 142996	regulatory protein [Bacillus subtilis]	67	41	492
856	1	1555	779	gi 780224	2K970.2 [Caenorhabditis elegans]	67	38	777
888	1	1614	850	gi 143715	77C start codon [Bacillus licheniformis]	67	40	765
1034	1	1190	597	gi 1205113	hypothetical protein (GB:U19201.15) [Haemophilus influenzae]	67	45	594
1062	1	636	319	gi 1303850	YqgC [Bacillus subtilis]	67	41	318
1067	1	918	460	pir[A32950]A329	probable reductase protein - Leishmania major	67	54	459

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
1358	1	3	293	gi 1001369	hypothetical protein (Synecocystis sp.)	67	44	291
2181	1	3	302	gi 1510416	hypothetical protein (SP-P31466) (Methanococcus jannaschii)	67	48	300
3000	1	1	507	gi 517205	67 kDa Myosin-crossreactive streptococcal antigen (Streptococcus pyogenes)	67	56	507
3066	1	464	234	gi 308861	GTC start codon (Lactococcus lactis)	67	46	231
3087	1	454	251	gi 1205366	oligopeptide transport ATP-binding protein (Haemophilus influenzae)	67	44	204
3101	1	2	256	gi 1531541	uroporphyrinogen III methyltransferase (Zea mays)	67	55	255
3598	1	728	393	gi 151259	HMG-CoA reductase (EC 1.1.1.88) (Pseudomonas mavalonitii) pif A44756 A44756 (hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.)	67	56	336
3765	2	584	366	gi 557489	nanD (Bacillus subtilis)	67	45	219
3788	1	658	398	pif S52915 S529	nitrate reductase alpha chain - Bacillus subtilis (fragment)	67	45	261
3883	1	2	265	gi 704397	cystathionine beta-lyase (Arabidopsis thaliana)	67	46	264
3926	1	2	340	gi 1463199	peptide-synthetase (Amycolatopsis mediterranei)	67	44	339
4417	1	82	396	gi 1205337	ribonucleotide transport ATP-binding protein (Haemophilus influenzae)	67	46	315
2	3	1075	3989	gi 535348	CodV (Bacillus subtilis)	66	42	918
15	6	2273	3542	gi 16491	SmbB (Synecococcus PCC7942)	66	37	270
31	9	8059	7826	gi 292046	luciferin (Homo sapiens)	66	44	234
31	10	9034	9258	gi 1204545	mercury scavenger protein (Haemophilus influenzae)	66	48	224
32	6	6347	5253	gi 998342	inducible nitric oxide synthase (Gallus gallus)	66	47	1095
44	13	8856	10124	gi 1510751	polydiphosphate cofactor biosynthesis moeA protein (Methanococcus jannaschii)	66	46	1269
48	2	1276	2868	gi 150209	ORF 1 (Mycoplasma mycoides)	66	40	1593
58	8	7178	8428	gi 665999	hypothetical protein (Bacillus subtilis)	66	47	1251
62	7	5143	4370	gi 1072398	pheD gene product (Rhizobium meliloti)	66	40	774
70	14	11693	10998	gi 809660	deoxyribose-phosphate aldolase (Bacillus subtilis) pif S49455 S49455 (deoxyribose-phosphate aldolase (EC 4.1.2.4) - acillus subtilis)	66	55	696
76	1	1	1305	gi 142440	ATP-dependent nuclease (Bacillus subtilis)	66	42	1305
91	6	9236	8205	gi 704397	cystathionine beta-lyase (Arabidopsis thaliana)	66	43	1032
102	5	3810	3265	gi 1204023	hypothetical protein (SP-P31805) (Haemophilus influenzae)	66	41	546

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Entry ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match name	% sim	% ident	Length (nt)
103	4	3418	2732	gi 571344	nitrate reductase gamma subunit [Bacillus subtilis] sp P42177 NARL_BACSU NITRATE REDUCTASE GAMMA CHAIN (EC 1.7.99.4); gi 1009369 Respiratory nitrate reductase [Bacillus subtilis] (SUB-160)	66	48	687
109	6	4243	4674	gi 170886	glucosamine-6-phosphate deaminase [Candida albicans] pir A46632 A46632 glucosamine-6-phosphate isomerase (EC 5.3.1.10) - aaat [Candida albicans]	66	45	432
112	17	17491	17712	gi 1323179	ORF YG111W [Saccharomyces cerevisiae]	66	33	222
116	2	4667	2637	gi 1491813	gamma-glutamyltranspeptidase [Bacillus subtilis]	66	43	2031
150	5	3189	2989	gi 1146224	putative [Bacillus subtilis]	66	30	201
172	5	3284	3662	gi 755152	highly hydrophobic integral membrane protein [Bacillus subtilis] sp P42953 TAGO_BACSU TAGOIC ACID TRANSLLOCATION PERMEASE PROTEIN AGG.	66	41	399
174	5	4592	3723	gi 1146241	pantothenate synthetase [Bacillus subtilis]	66	49	870
175	4	3209	2880	gi 642655	unknown [Rhizobium meliloti]	66	29	330
175	11	8743	7994	gi 854655	Na/H antiporter system [Bacillus alcalophilus]	66	43	750
190	5	7079	5727	gi 451072	di-tripeptide transporter [Lactococcus lactis]	66	40	1353
195	15	13919	13713	gi 1322411	unknown [Mycobacterium tuberculosis]	66	42	207
217	3	2822	2595	gi 1143542	alternative stop codon [Hattus norvegicus]	66	36	228
217	9	7133	6135	gi 1458327	P08F3_4 gene product [Caenorhabditis elegans]	66	47	999
218	3	43	1041	gi 409541	CBRA protein [Erwinia chrysanthemi]	66	42	999
241	3	2102	1053	gi 153067	peptidoglycan hydrolase [Staphylococcus aureus]	66	53	1050
261	3	1178	646	gi 1510859	M. jannaschii predicted coding region M30790 [Methanococcus jannaschii]	66	40	531
263	3	3731	2973	gi 1205865	tetrahydrodipicolinate N-succinyltransferase [Haemophilus influenzae]	66	47	759
272	8	6548	5484	gi 882101	high affinity nickel transporter [Alcaligenes eutrophus] sp P23516 HOXA_ALCEU HIGH-AFFINITY NICKEL TRANSPORT PROTEIN	66	44	1065
276	3	2805	2104	gi 1208965	hypothetical 23.3 kd protein [Escherichia coli]	66	47	702
278	2	2810	1784	gi 1488662	phosphatase-associated protein [Bacillus subtilis]	66	48	1047
278	3	3830	2952	gi 303560	ORE27; [Escherichia coli]	66	45	879
279	2	3894	2218	gi 1185289	2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase [Bacillus subtilis]	66	48	1677
288	4	2535	2275	gi 1256625	putative [Bacillus subtilis]	66	42	261
292	2	1333	942	gi 1511604	M. jannaschii predicted coding region M31651 [Methanococcus jannaschii]	66	30	192

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
294	1	1116	gi 216314	esterase [Bacillus stearothermophilus]	66	45	558
297	4	2913	gi 954794	cytochrome a assembly facto [Bacillus subtilis] sp P24009 COXX_BACSU PROBABLE CYTOCHROME C OXIDASE ASSEMBLY FACTOR.	66	45	936
316	4	2053	gi 1107839	alginate lyase [Pseudomonas aeruginosa]	66	40	610
338	4	2460	gi 520750	blotol synthetase [Bacillus sphaericus]	66	59	159
339	1	1214	gi 467468	7, 8-dihydro-6-hydroxymethylpterin-pyrophosphokinase [Bacillus ubtilis]	66	52	480
363	1	863	gi 581649	epiC gene product [Staphylococcus epidermidis]	66	47	861
366	2	232	gi 1103505	unknown [Schizosaccharomyces pombe]	66	53	252
367	4	2468	sp P20692 TYRA	PREPHENATE DEHYDROGENASE (EC 1.3.1.12) (PDH)	66	50	624
372	3	2150	gi 467416	unknown [Bacillus subtilis]	66	38	552
378	1	212	gi 147109	purine nucleoside phosphorylase [Escherichia coli]	66	50	798
401	1	462	gi 388263	p-aminobenzoic acid synthase [Streptomyces griseus] pir JN0531 JN0531 p-aminobenzoic acid synthase - Streptomyces riseus	66	46	462
404	7	4826	gi 406744	cytidine deaminase [Bacillus subtilis]	66	51	429
411	2	1738	gi 1460081	unknown [Mycobacterium tuberculosis]	66	44	636
420	1	2	gi 1046024	Na+ ATPase subunit J [Mycoplasma genitalium]	66	49	540
431	1	858	gi 3500008	M. jannaschii predicted coding region MJ1154 [Methanococcus jannaschii]	66	50	858
443	7	5679	gi 852076	MrgA [Bacillus subtilis]	66	46	381
444	3	3405	gi 153047	lysoestaphin (ttg start codon) [Staphylococcus simulans] pir A25881 A25881 lysoestaphin precursor - Staphylococcus simulans ap P10547 LSTP_STASI LYOSTAPHIN PRECURSOR (EC 3.5.1.-)	66	51	993
561	1	956	gi 1204905	DNA-3-methyladenine glycosylase I [Haemophilus influenzae]	66	45	477
562	3	1066	gi 1046082	M. genitalium predicted coding region M0372 [Mycoplasma genitalium]	66	52	318
576	3	11	gi 305014	ORF_0234 [Escherichia coli]	66	43	714
577	3	1190	gi 1001353	hypothetical protein [Synachocystis sp.]	66	52	288
584	1	2	gi P24204 YEB	LYSOTHEICAL 46.7 KD PROTEIN IN ASBB-RUVB INTERGENIC REGION (ORFU)	66	48	330
592	1	1410	gi 928839	ORF266: putative [Lactococcus lactis phase BK5-T]	66	51	705
601	1	1433	gi 1488695	novel antigen: orf-2 [Staphylococcus aureus]	66	55	716

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
619	3	468	845	gi1746573	similar to H. musculus transport system membrane protein, Ncramp PIR:A40739 and S. cerevisiae SHF1 protein (PIR:A45154) <i>Caenorhabditis elegans</i>	66	45	378
706	2	561	355	gi1804808	unknown protein [Rattus norvegicus]	66	46	207
734	2	673	512	gi1519085	phosphatidylcholine binding immunoglobulin heavy chain IGH variable region [Mus musculus]	66	60	162
740	1	3	317	gi11209272	argininosuccinate lyase [Campylobacter jejuni]	66	47	315
764	1	310	747	gi1435296	alkaline phosphatase like protein [Lactococcus lactis] pir S19339 S19339 alkaline phosphatase-like protein - Lactococcus actis	66	42	418
852	1	338	171	gi1536955	CG Site No. 361 [Escherichia coli]	66	43	168
886	1	3	158	gi1289272	ferrichrome-binding protein [Bacillus subtilis]	66	44	156
889	1	462	232	gi1833061	HCMVUL77 (AA 1-642) [Human cytomegalovirus]	66	66	231
893	1	2	247	gi1149008	putative [Helicobacter pylori]	66	45	246
900	1	1425	733	gi1580842	P3 [Bacillus subtilis]	66	51	693
906	2	2300	1473	gi1790945	eryl-alcohol dehydrogenase [Bacillus subtilis]	66	53	828
947	1	79	549	gi1410117	diaminopimelate decarboxylase [Bacillus subtilis]	66	47	471
950	1	1100	552	gi148713	orf145 [Staphylococcus aureus]	66	35	549
955	2	89	475	gi11204390	uridine kinase (uridine monophosphokinase) [Haemophilus influenzae]	66	50	387
981	2	1304	997	gi1457146	phoptry protein [Plasmodium yoelii]	66	10	117
986	1	25	315	gi1305002	ORF_1356 [Escherichia coli]	66	31	291
1057	1	3	203	gi11031853	YggP [Bacillus subtilis]	66	40	201
1087	1	1	294	gi1575913	unknown [Saccharomyces cerevisiae]	66	53	294
1105	1	1	231	gi11045799	methylgalactoside permease ATP-binding protein [Mycoplasma genitalium]	66	46	231
1128	1	2	574	gi11001493	hypothetical protein [Synchocystis sp.]	66	46	573
1150	1	498	250	gi11499034	M. jannaschii predicted coding region M0255 [Methanococcus jannaschii]	66	40	249
1180	2	707	453	gi1215908	DNA polymerase (g43) [Bacteriophage T4]	66	46	255
1204	1	1123	587	gi11256653	DNA-binding protein [Bacillus subtilis]	66	58	537
1342	1	1	402	gi11208474	hypothetical protein [Synchocystis sp.]	66	53	402
1761	2	589	398	gi1215811	cell fiber protein [Bacteriophage T3]	66	50	192

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
1983	1	499	251	gi1045935	DNA helicase II (Mycoplasma genitalium)	66	40	249
2103	2	176	400	gi1929798	Precursor for the major microtoe surface antigens (Plasmodium alciptarum)	66	46	225
2341	1	973	188	gi1256623	lexodeoxyribonuclease (Bacillus subtilis)	66	38	186
2458	1	325	164	gi1019410	unknown (Schizosaccharomyces pombe)	66	47	162
2505	1	468	235	gi1510184	putative transcriptional regulator (Methanococcus jannaschii)	66	39	234
2525	1	558	280	gi1000695	cytotoxin L (Clostridium sordellii)	66	44	279
2915	1	1	275	gi1765073	autolysin (Staphylococcus aureus)	66	47	273
3005	1	114	305	gi1205784	heterocyst maturation protein (Haemophilus influenzae)	66	46	192
3048	1	80	277	gi1303813	Yqew (Bacillus subtilis)	66	42	198
3071	1	1	189	gi1070014	protein-dependent (Bacillus subtilis)	66	41	189
3081	1	404	225	gi1984212	unknown (Schizosaccharomyces pombe)	66	44	180
3090	2	580	386	gi1204987	DNA polymerase III, alpha chain (Haemophilus influenzae)	66	48	195
3314	1	1	387	gi1009366	Respiratory nitrate reductase (Bacillus subtilis)	66	49	387
3739	1	796	400	gi1109684	ProV (Bacillus subtilis)	66	47	399
3786	1	402	202	gi1853760	acyl-CoA dehydrogenase (Bacillus subtilis)	66	60	201
3974	1	590	347	gi1263952	glucuronate permease (Bacillus licheniformis)	66	46	249
4240	1	1	350	gi151259	HMG-CoA reductase (EC 1.1.1.88) (Pseudomonas mevalonii) pir[A44/56/A44/56] hydroxymethylglucaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	66	51	138
4604	1	7	234	pir[A26/13]BHC	hemocyanin subunit II - Atlantic horseshoe crab	66	46	228
4	9	8845	9750	gi145646	cynR (Escherichia coli)	65	35	906
6	5	2708	3565	gi1887824	ORF_0310 (Escherichia coli)	65	47	858
13	1	1993	998	gi143402	recombination protein (ttg start codon) (Bacillus subtilis) gi1303923 RecN (Bacillus subtilis)	65	44	996
15	7	2493	3524	gi1403126	cspD gene product (Alcaligenes eutrophus)	65	38	1032
18	3	1908	1372	gi1349187	acyltransferase (Saccharomyces cerevisiae)	65	50	537
21	3	1467	2492	gi1449518	phosphoribosyl anthranilate transferase (Lactococcus lactis) (pir[S35126/S35126 anthranilate phosphoribosyltransferase (EC 4.2.1.18) - Lactococcus lactis subsp. lactis	65	52	1026
25	4	3374	4312	gi1502420	malonyl-CoA:Acyl carrier protein transacylase (Bacillus subtilis)	65	44	939

TABLE 2



S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
27	2	390	626	gi 1212729	lqhJ [Bacillus subtilis]	65	45	237
31	12	11040	10387	gi 509245	D-hydroxyacetate dehydrogenase [Lactobacillus delbrueckii]	65	41	654
38	24	19172	19528	gi 547	H-protein [Flavaria croquisi]	65	41	357
44	2	790	1746	gi 405882	yeK [Escherichia coli]	65	46	957
44	12	9356	8832	gi 1205905	molybdenum cofactor biosynthesis protein [Haemophilus influenzae]	65	50	525
45	8	6635	7588	gi 493074	ApbA protein [Salmonella typhimurium]	65	46	934
51	2	580	1503	gi 580897	OppA gene product [Bacillus subtilis]	65	45	924
52	1	225	953	gi 1205518	NAD(P)H-flavin oxidoreductase [Haemophilus influenzae]	65	45	729
55	4	1339	1058	pir A4459 A444	troponin T beta Tm-5 - rabbit	65	41	282
67	9	7421	8272	gi 143607	sporulation protein [Bacillus subtilis]	65	42	852
73	5	4446	5375	gi 1204896	lysophospholipase L2 [Haemophilus influenzae]	65	37	930
74	1	954	478	gi 1204844	H. influenzae predicted coding region H10594 [Haemophilus influenzae]	65	50	477
77	1	2	757	gi 1046082	H. genitalium predicted coding region MG372 [Mycoplasma genitalium]	65	46	756
77	2	795	1433	gi 1222116	permease [Haemophilus influenzae]	65	37	639
81	3	4728	3454	gi 1001708	hypothetical protein [Synechocystis sp.]	65	49	1275
91	7	8548	8357	gi 1399263	cystathionine beta-lyase [Emeritella nidulans]	65	40	192
98	3	1608	1988	gi 467423	unknown [Bacillus subtilis]	65	38	381
98	4	2250	2987	gi 467424	unknown [Bacillus subtilis]	65	45	738
102	3	2598	2119	gi 1511532	N-terminal acetyltransferase complex, subunit ARO1 [Methanococcus jannaschii]	65	39	480
102	4	3647	2862	gi 1204637	H. influenzae predicted coding region H10388 [Haemophilus influenzae]	65	32	786
103	9	110851	9841	gi 142695	[S-adenosyl]-L-methionine:uroporphyrinogen III methyltransferase [Bacillus megaterium]	65	47	1011
103	10	10439	10119	gi 1710021	nitrite reductase (nirD) [Bacillus subtilis]	65	51	321
106	2	262	1140	gi 39881	ORF 311 (AA 1-311) [Bacillus subtilis]	65	44	879
109	5	3909	4268	gi 1204399	glucosamine-6-phosphate deaminase protein [Haemophilus influenzae]	65	44	360
109	10	7165	8595	gi 536955	CG Site No. 361 [Escherichia coli]	65	41	1431

TABLE 2

3 aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
110	4	3688	3915	[gi1407881]	stringent response-like protein [Streptococcus equisimilis] pir[S3975]S3975 stringent response-like protein - Streptococcus quiesimilis	65	45	228
110	5	3882	4295	[gi1407880]	[ORF1] Streptococcus equisimilis	65	50	414
110	6	4231	4380	[gi1139574]	[orf2] Streptomyces griseus	65	56	150
112	10	9218	8640	[gi11204571]	[H] influenzae predicted coding region H10318 [Haemophilus influenzae]	65	52	573
112	12	12049	11288	[gi1710496]	[transcriptional activator protein [Bacillus brevis]	65	32	762
125	1	2	202	[gi11151158]	[repeat organellar protein [Plasmodium chabaudi]	65	39	201
126	1	3	422	[gi137589]	[precursor [Homo sapiens]	65	46	420
127	11	10713	12658	[gi11064809]	[homologous to sp:UTRA_EC01 [Bacillus subtilis]	65	41	1926
143	8	7543	7004	[gi1216513]	[mutator mutT [AT-CC transversion] [Escherichia coli]	65	56	540
145	5	3587	3818	[gi11209768]	[D02_orf569 [Mycoplasma pneumoniae]	65	27	252
150	4	3442	2841	[gi11146225]	[putative [Bacillus subtilis]	65	37	642
166	1	3858	1948	[g1148104]	[beta-1,4-N-acetylmuramylhydrolase [Enterococcus hirae] pir[A42296]A42296 lysosyme 2 [EC 3.2.1.-] precursor - Enterococcus irae (ATCC 9790)	65	50	1911
188	6	3195	4178	[gi1151943]	[ORF3; putative [Rhodobacter capsulatus]	65	46	984
189	9	4392	4785	[gi158812]	[ORF IV [AA 1-480] [Fujimori mosaic virus]	65	40	178
195	6	7900	5272	[gi1145220]	[alanyl-tRNA synthetase [Escherichia coli]	65	44	2637
195	7	11059	8104	[gi1882711]	[exonuclease V alpha-subunit [Escherichia coli]	65	38	2486
206	16	11696	18191	[g1408115]	[ornithine acetyltransferase [Bacillus subtilis]	65	53	1296
217	4	3844	3215	[gi11205974]	[5'guanylate kinase [Haemophilus influenzae]	65	41	630
220	4	5265	3751	[gi1580920]	[rodd [graa] polypeptide [AA 1-673] [Bacillus subtilis] pir[S06048]S06048 proteobactin [Bacillus subtilis] sp[P13484]TAGP BACSU PROBACTIN POLYGLYCEROL-PHOSPHATE LPH-GLYCOSYLTRANSFERASE [EC 2.4.1.52] [TECHIOIC ACID BIOSYNTHESIS ROUTIN E]	65	40	1515
236	5	2127	3709	[gi11146200]	[DNA or RNA helicase, RNA-dependent ATPase [Bacillus subtilis]	65	46	1381
237	3	1902	2513	[gi1149379]	[HlsB [Lactococcus lactis]	65	46	612
241	4	4968	4195	[gi11205308]	[ribonuclease III [EC 3.1.264] [RNASE III] [Haemophilus influenzae]	65	50	774
252	1	1278	940	[gi11204989]	[hypothetical protein [GB U00022.9] [Haemophilus influenzae]	65	40	319
261	5	4780	3794	[g1145527]	[fecD [Escherichia coli]	65	43	987

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match %	Match accession	Match gene name	% sim	% ident	Length (nt)
274	1	3	278	94	gi1496558	orfX (Bacillus subtilis)	65	42	276
301	2	982	815	93	gi152418	unknown (Bacillus subtilis)	65	45	168
307	4	3586	2864	94	gi1070014	protein-dependent (Bacillus subtilis)	65	40	723
335	2	2286	1399	91	gi1146913	N-acetylglucosamine transport protein (Escherichia coli) p1rB29895 MOEC2N phosphotransferase system enzyme 11 (EC 7.1.69), N-acetylglucosamine-specific - Escherichia coli sp1P03223 PTMA_ECOLI PTS SYSTEM, N-ACETYLGLUCOSAMINE-SPECIFIC IIABC COMPONENT (EIIA)	65	50	888
338	5	4320	3170	93	gi11277029	biotin synthase (Bacillus subtilis)	65	49	951
343	3	1490	2800	93	gi1143264	membrane-associated protein (Bacillus subtilis)	65	48	1311
344	4	2781	2531	93	gi11050540	tRNA-glutamine synthetase (Lupinus luteus)	65	34	231
358	3	3421	3621	93	gi11146220	NAD+ dependent glycerol-3-phosphate dehydrogenase (Bacillus subtilis)	65	47	201
364	1	238	699	93	gi1140128	ORF1 (Staphylococcus aureus)	65	51	462
379	1	1	576	93	gi1143331	alkaline phosphatase regulatory protein (Bacillus subtilis) p1rA27650 A27650 regulatory protein phoR - Bacillus subtilis sp1P23545 PHOR_BACSU ALKALINE PHOSPHATASE SYNTHESIS SENSOR PROTEIN PHOR (EC 2.7.3.-)	65	40	576
379	3	3666	4346	93	gi1143268	dihydrolipoamide transsuccinylase (odhB; EC 2.3.1.61) (Bacillus subtilis)	65	50	681
428	1	187	483	93	gi11420465	ORF YOR195W (Saccharomyces cerevisiae)	65	45	297
438	2	272	838	93	gi1143498	degS protein (Bacillus subtilis)	65	38	567
444	11	9280	10215	93	gi11204756	ribokinase (Haemophilus influenzae)	65	47	936
449	2	1241	1531	93	gi1599848	Na/H antiporter homolog (Lactococcus lactis)	65	41	291
478	2	1452	865	93	gi11045942	glycyl-tRNA synthetase (Hycoplasma genitalium)	65	39	588
479	1	1032	517	93	gi11498192	putative (Pseudomonas aeruginosa)	65	40	516
480	6	4312	5637	93	gi1415662	UDP-N-acetylglucosamine 1-carboxyvinyl transferase (Acinetobacter baumannii)	65	48	1326
484	1	2	430	93	gi1146751	transmembrane protein (KdpD) (Escherichia coli)	65	44	429
499	1	54	932	93	gi1603456	reductase (Leishmania major)	65	53	879
505	1	914	459	93	gi11518853	IoafA (Salmonella typhimurium)	65	39	456
571	2	1509	883	93	gi149399	open reading frame upstream glfE (Escherichia coli) ir S37754 S37754 hypothetical protein XE (glfE 5' region) - Escherichia coli	65	44	627
611	2	506	270	93	gi110961	RAP-2 (Plasmodium falciparum)	65	40	237

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
705	1	564	283	[gi710020]	nitrate reductase (nirB) [Bacillus subtilis]	65	52	282
712	1	1	177	[gi289272]	ferrichrome-binding protein [Bacillus subtilis]	65	37	177
712	2	196	354	[gi289272]	ferrichrome-binding protein [Bacillus subtilis]	65	17	159
743	1	2	631	[gi310631]	ATP binding protein [Streptococcus gordonii]	65	45	610
749	2	393	779	[gi167374]	single strand DNA binding protein [Bacillus subtilis]	65	29	387
762	1	1698	850	[gi160399]	multidrug resistance protein [Plesiomonas falciparum]	65	48	849
788	1	85	315	[gi1129096]	unknown protein [Bacillus sp.]	65	35	231
850	1	1	408	[gi1006604]	hypothetical protein [Synecocystis sp.]	65	37	408
908	1	1	444	[gi1199546]	2362 [Saccharomyces cerevisiae]	65	46	444
925	1	1	174	[gi1256653]	DNA-binding protein [Bacillus subtilis]	65	54	174
1031	1	26	212	[gi238657]	AppCyclochrome d oxidase, subunit 1 homolog [Escherichia coli, K12, epitide, 514 aa]	65	47	207
1037	1	414	262	[gi1491813]	gamma-glutamyltranspeptidase [Bacillus subtilis]	65	46	153
1053	1	344	175	[gi642655]	unknown [Rhizobium meliloti]	65	34	174
1149	1	1399	752	[gi1162980]	ribulose-5-phosphate 3-epimerase [Spinacia oleracea]	65	48	648
1214	1	881	495	[gi1205959]	lactam utilization protein [Haemophilus influenzae]	65	45	387
1276	1	476	276	[gi1535493]	site-specific DNA-methyltransferase Stel (EC 2.1.1.-) - Streptococcus sanguis	65	35	201
1276	2	900	577	[gi1473794]	"ORF" [Escherichia coli]	65	34	324
2057	1	272	136	[gi1635699]	TrsH [Yersinia enterocolitica]	65	21	135
2521	1	336	169	[gi1045789]	hypothetical protein (CBU4001.76) [Mycoplasma genitalium]	65	41	168
2974	1	590	297	[gi152052]	enantiomerase selective amidase [Rhodococcus sp.]	65	45	294
3031	1	306	154	[gi1201024]	hypothetical JOK protein (tmP140 5' region) - fruit fly (Drosophila melanogaster)	65	35	151
3069	1	3	278	[gi144906]	product homologous to E. coli thioredoxin reductase, J Biol Chem, 1988; 263:9015-9019, and to P52a protein of alkyl hydroperoxide oxidase from S. typhimurium; J Biol Chem, (1990) 265:10535-10540, pan reading frame A [Clostridium pasteurianum]	65	46	276
3146	1	282	142	[gi149315]	[ORF] gene product [Bacillus subtilis]	65	47	141
3170	1	679	341	[gi11507711]	indolepyruvate decarboxylase [Erwinia herbicola]	65	44	339

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
3546	1	1	103	gi1450688	hsdM gene of E. coli product [Escherichia coli] pir[S18437/S38437] hsdM protein - Escherichia coli pir[S09629/S09629] hypothetical protein A - Escherichia coli (SUB 40-520)	65	42	303
3782	1	2	328	gi166412	NADH-glutamate synthase (Medicago sativa)	65	42	327
3990	1	374	189	gi1009366	Respiratory nitrate reductase [Bacillus subtilis]	65	53	186
4012	1	613	308	gi1132127	ORP YGR087c [Saccharomyces cerevisiae]	65	50	306
4278	2	726	364	gi1197667	vitellogenin [Anolis pulchellus]	65	42	363
19	4	4259	5518	gi145727	dead [Escherichia coli]	64	45	1260
19	6	7639	6926	gi1016232	lyc27 gene product [Cyanophora paradoxa]	64	36	714
20	8	7053	6454	gi1765073	autolysin [Staphylococcus aureus]	64	47	600
31	13	12706	11537	gi1414009	lipo-85d gene product [Bacillus subtilis]	64	45	1170
33	4	2388	4364	gi1204696	fructose-permease IIBC component [Haemophilus influenzae]	64	47	1977
36	3	1871	3013	gi1290503	glutamate permease [Escherichia coli]	64	40	1143
37	6	4065	4409	gi139815	orf 2 gene product [Bacillus subtilis]	64	46	345
45	9	7852	8760	gi1230585	nucleotide sugar epimerase [Vibrio cholerae O139]	64	53	909
53	3	1540	1899	gi1303961	lycJ [Bacillus subtilis]	64	50	360
56	6	4793	3855	gi1457514	glcC [Bacillus subtilis]	64	45	919
56	24	30002	30247	gi1470331	similar to zinc fingers [Caenorhabditis elegans]	64	42	246
62	4	2759	2421	gi1642655	unknown [Rhizobium meliloti]	64	28	339
85	6	7178	6027	gi1457702	5-aminimidazole ribonucleotide-carboxylase [Pichia methanolica] pir[S5912/S5912] phosphoribosylaminoimidazole carboxylase (EC 1.1.21) - yeast [Pichia methanolica]	64	46	1152
96	9	9251	10030	gi1511513	ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii]	64	42	780
100	1	1	600	gi1765073	autolysin [Staphylococcus aureus]	64	44	600
106	5	3868	4854	gi1466778	lysine specific perase [Escherichia coli]	64	46	987
123	2	838	554	gi1467484	unknown [Bacillus subtilis]	64	47	285
127	8	7514	7810	gi1710061	isotype-specific antigen [African horse sickness virus] pir[S27891/S27891] capsid protein VP2 - African horse sickness virus	64	28	297
131	7	7134	6721	gi1511160	M. jannaschii predicted coding region M1163 [Methanococcus jannaschii]	64	46	414

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	A sim	A ident	length (nt)
142	5	3465	4817	gi 1173517	riboflavin synthase alpha subunit [Actinobacillus pleuropneumoniae]	64	44	639
143	1	709	356	gi A32950 A329	probable reductase protein - Leishmania major	64	52	354
149	10	3555	3295	gi 398151	major surface antigen MS2 [Pneumocystis carinii]	64	44	261
154	4	3134	2307	gi 3984587	DinP [Escherichia coli]	64	50	828
161	5	1855	4880	gi 3933304	ORF72 [Bacillus subtilis]	64	37	1026
165	1	33	791	gi 467483	unknown [Bacillus subtilis]	64	38	759
175	6	6355	4214	gi 1072398	phd gene product [Rhizobium meliloti]	64	42	1512
188	3	2042	2500	gi 1001961	MHC class II analog [Staphylococcus aureus]	64	45	459
195	14	13667	13446	gi 396380	No definition line found [Escherichia coli]	64	47	222
206	15	16429	16938	gi 304134	argC [Bacillus stearothermophilus]	64	49	510
215	1	560	282	gi 142359	ORF 6 [Acetobacter vinelandii]	64	39	279
243	7	7818	6928	gi 414014	spe-90d gene product [Bacillus subtilis]	64	49	891
258	2	1330	845	gi 664754	PL7 [Listeria monocytogenes]	64	38	486
259	1	462	232	gi 1459663	M. jannaschii predicted coding region M30817 [Methanococcus jannaschii]	64	52	231
263	6	6565	5567	gi 142828	aspartate semialdehyde dehydrogenase [Bacillus subtilis] sp Q04797 DHAS-DACSU ASPARTATE-SERIAL-DEHYDRO-DEHYDROGENASE (EC 2.1.11) (ASA DEHYDROGENASE)	64	48	999
271	1	3	1163	gi 467091	hflX; B2235_C2_202 [Mycobacterium leprae]	64	44	1161
280	1	173	1450	gi 1301839	YqfR [Bacillus subtilis]	64	43	1278
293	1	2532	1267	gi 147345	primosomal protein n' [Escherichia coli]	64	45	1266
295	2	742	1488	gi 459266	Potential membrane spanning protein [Staphylococcus hominis] pir S42932 S42932 potential membrane spanning protein - Staphylococcus hominis	64	39	747
301	5	1625	1446	gi 580835	lysine decarboxylase [Bacillus subtilis]	64	35	180
315	4	5064	3949	gi 141396	guinol oxidase [Bacillus subtilis]	64	45	1116
321	1	1264	635	gi 710496	transcriptional activator protein [Bacillus brevis]	64	41	630
333	5	4520	4239	gi 3314295	ORF2; putative 19 kDa protein [Listeria monocytogenes]	64	43	282
342	1	1	549	gi 142940	ftsA [Bacillus subtilis]	64	38	549
353	3	2878	2324	gi 537049	ORF_0470 [Escherichia coli]	64	44	555

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
373	2	827	1658	gi 525295 A128	oxoglutarate dehydrogenase (liponate) (SC 1.2.4.2) - Bacillus subtilis	64	47	2832
404	6	4429	4839	pir A36933 A369	diacylglycerol kinase homolog - Streptococcus mutans	64	35	411
407	1	2020	1133	gi 969026	OrfX [Bacillus subtilis]	64	41	888
425	1	1109	591	gi 1146177	phosphotransferase system glucose-specific enzyme II [Bacillus subtilis]	64	44	519
443	6	4082	4798	gi 147109	purine nucleoside phosphorylase [Escherichia coli]	64	51	717
450	2	1035	1604	gi 106376	ORF_0162 [Escherichia coli]	64	38	570
470	5	1680	6107	gi 1369948	host interacting protein [Bacteriophage B1]	64	45	4428
486	4	1911	1471	gi 1205582	spermidine/putrescine transport system permease protein [Haemophilus influenzae]	61	35	441
497	1	2217	1159	sp P36929 FHU_E	FHU PROTEIN	64	38	1059
501	1	3	410	gi 142450	ehcC protein [Bacillus subtilis]	64	38	408
514	1	3	290	gi 1204496	H. influenzae predicted coding region HI0238 [Haemophilus influenzae]	64	34	288
551	4	3162	3323	gi 1204511	bacterioferritin conjugatory protein [Haemophilus influenzae]	64	41	162
603	4	759	956	gi 755823	NADH dehydrogenase F [Streptococcus amarae]	64	35	198
653	2	940	746	gi 1212234	dicarboxylic amino acids Dipeptidase [Saccharomyces cerevisiae]	64	41	195
660	3	3801	2257	sp P46133 YHAI_	HYPOTHETICAL PROTEIN IN OUT 5 REGION (FRAGMENT)	64	39	1545
695	1	11	502	gi 1001383	hypothetical protein [Synecocystis sp.]	64	41	492
702	1	3	752	gi 142865	DNA primase [Bacillus subtilis]	64	46	750
826	1	1	339	gi 971336	arginyl tRNA synthetase [Bacillus subtilis]	64	50	339
838	1	1831	917	gi 1254775	pfoS/R [Treponema pallidum]	64	41	915
864	3	675	944	gi 39833	cyclomaltodextrin glucanotransferase [Bacillus stearothermophilus] i 39835	64	47	270
887	1	3	677	gi 153002	cyclomaltodextrin glucanotransferase [Bacillus stearothermophilus]	64	46	675
924	2	1172	963	gi 111976	enterotoxin type E precursor [Staphylococcus aureus] pir A28179 A28179	64	46	675
1049	2	800	606	gi 1049115	enterotoxin E precursor - Staphylococcus aureus sp P12993 ETX_E STAU	64	42	195
1067	2	999	748	gi 1151072	ENTEROTOXIN TYPE E PRECURSOR (SEE)	64	50	252
924	2	1172	963	gi 111976	fibrinogen-binding protein [Staphylococcus aureus] pir S34270 S34270	64	41	210
1049	2	800	606	gi 1049115	fibrinogen-binding protein - Staphylococcus aureus	64	42	195
1067	2	999	748	gi 1151072	Pap60 [Bacillus subtilis]	64	42	195
1067	2	999	748	gi 1151072	Hda precursor [Haemophilus ducreyi]	64	50	252

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
1120	1	50	202	gi 42439	ATP-dependent nuclease [Bacillus subtilis]	64	30	153
1125	1	751	377	gi 581648	epiB gene product [Staphylococcus epidermidis]	64	44	375
1508	1	402	214	gi A01365 TVMS	transforming protein K-12a - mouse	64	47	189
2472	1	2	358	gi 487282	Na <sup>+</sup> -ATPase subunit J [Enterococcus hirae]	64	36	357
2989	1	520	356	gi 304134	argC [Bacillus stearothermophilus]	64	50	165
3013	1	630	352	gi 551699	cytochrome oxidase subunit I [Bacillus firmus]	64	51	279
3034	1	546	274	gi 1204349	hypothetical protein (GB:GB D90212.3) [Haemophilus influenzae]	64	50	273
3197	1	613	308	gi 1009366	Respiratory nitrate reductase [Bacillus subtilis]	64	46	306
3303	1	90	362	gi 1107839	alginate lyase [Pseudomonas aeruginosa]	64	43	271
3852	2	82	288	gi 216746	D-lactate dehydrogenase [Lactobacillus plantarum]	64	42	207
3868	1	1	312	gi 149415	putative [Lactococcus lactis]	64	48	312
3918	1	660	331	gi 5532	acetyl-CoA acyltransferase [Vibrio lipolytica]	64	46	330
4000	1	112	378	gi 944688	unknown [Saccharomyces cerevisiae]	64	44	267
4009	1	81	368	gi 39372	grab gene product [Bacillus brevis]	64	41	288
4166	1	2	349	gi 149435	putative [Lactococcus lactis]	64	46	348
4366	1	2	307	gi 216267	OMP2 [Bacillus megaterium]	64	41	306
4457	1	2	400	gi 1197657	vitallogenin [Anolis pulchellus]	64	43	399
11	3	1539	2438	gi 438228	OMP C [Staphylococcus aureus]	63	32	900
24	7	5611	5423	gi 1369943	a1 gene product [Bacteriophage B1]	63	34	189
29	1	1	390	gi 467441	expressed at the end of exponential growth under conditions in which the enzymes of the TCA cycle are repressed [Bacillus subtilis] gi 467441	63	43	390
31	6	6329	5712	gi 496943	ORF [Saccharomyces cerevisiae]	63	47	618
44	123	114669	15019	gi A04446 QDEC	hypothetical protein F-92 - Escherichia coli	63	36	351
48	6	4403	6250	gi 43498	pyruvate synthase [Halobacterium halobium]	63	42	1848
50	5	3869	4738	gi 413967	ipa-43d gene product [Bacillus subtilis]	63	43	870
53	6	6764	5742	gi 474176	regulator protein [Staphylococcus xylosus]	63	49	1023

TABLE 2



TABLE 2 - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match	Accession	Match gene name	λ nm	t ident	length (nt)
56	14	15880	17607	91	167409	DNA polymerase III subunit [Bacillus subtilis]	63	44	1728
57	11	7945	7376	91	137036	ORF_0158 [Escherichia coli]	63	39	570
62	3	2479	2114	91	142656	unknown [Rhizobium meliloti]	63	41	365
70	8	6562	7353	91	1399821	PhoC [Rhizobium meliloti]	63	46	792
75	2	223	927	91	149376	Hsd [Lactococcus lactis]	63	45	705
78	5	4912	4403	91	141950	ipa-26d gene product [Bacillus subtilis]	63	42	510
91	5	9076	7220	91	1466997	metH2, B2126-CL157 [Mycobacterium leprae]	63	41	1857
91	8	10566	9448	91	1204344	cystathionine gamma-synthase [Haemophilus influenzae]	63	45	1119
120	1	21	1508	91	1882657	sulfite reductase (NADPH) [flavoprotein beta subunit [Escherichia coli]]	63	46	1444
120	4	2723	4125	91	1665994	hypothetical protein [Bacillus subtilis]	63	34	1404
127	7	6064	7566	91	140162	murE gene product [Bacillus subtilis]	63	44	1503
149	6	2321	2106	91	148503	dnaK [Erysipelothrix rhusiopathiae]	63	40	216
149	26	10445	10170	91	14870	ORF 2, has similarity to DNA polymerase [Saccharomyces kluyveri] r[S15961/S15961] hypothetical protein 2 - yeast [Saccharomyces yverii] plasmid pSK1	63	42	276
164	2	507	1298	91	135476	CDP-diglyceride synthetase [Escherichia coli]	63	44	792
166	6	7009	8164	91	151912	fructose enzyme II [Rhodospirillum rubrum]	63	31	1746
169	4	1704	1886	91	152886	elongation factor Ts (tsf) [Spiroplasma citri]	63	48	183
188	5	3145	2951	91	1334547	GlyCOT 114 gtp IB protein [Podospora anserina]	63	42	195
195	13	11767	112804	91	1606100	ORF_0335 [Escherichia coli]	63	40	1034
201	2	607	2283	91	1433534	arginyl-tRNA synthetase [Corynebacterium glutamicum] pir[A49936/A49936] arginine-tRNA ligase [EC 6.1.1.19] - oryzae [Corynebacterium glutamicum]	63	46	1677
206	14	15893	116489	91	1380828	N-acetyl-glutamate-gamma-semialdehyde dehydrogenase [Bacillus subtilis]	63	49	597
220	5	7769	5766	91	1216334	seca protein [Bacillus subtilis]	63	42	2004
221	1	74	907	91	1677945	AppA [Bacillus subtilis]	63	42	831
227	3	944	1708	91	1510558	cohyric acid synthase [Methanococcus jannaschii]	63	46	765
261	2	804	1070	91	1486511	ORF YKR054c [Saccharomyces cerevisiae]	63	45	267
269	2	1606	1960	91	148221	DNA-dependent ATPase, DNA helicase [Escherichia coli] pir[J50137/BVECRQ] recQ protein - Escherichia coli	63	42	1647

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
278	1	7417	gi 699273	cystathionine gamma-synthase [Mycobacterium leprae] sp P46807 METB_MYCLE CYSTATHIONINE GAMMA-SYNTASE [EC 4.2.99.9] O-SUCCINYLMOMOSERINE (THIOL)- LYASE	63	41	1242
287	2	738	gi 405133	putative [Bacillus subtilis]	63	38	996
295	1	748	gi 1239983	hypothetical protein [Bacillus subtilis]	63	41	747
328	3	2148	gi 45302	carrier protein (AA : - 47) [Pseudomonas aeruginosa] ir S11497 S11497 branched-chain amino acid ti - sport protein brab - eudomonas aeruginosa	63	36	987
362	2	1526	sp P51316 SERA_	D-3-PHOSPHOGLYCERATE DEHYDROGENASE [EC 1.1.1.95] (PUDH)	63	38	411
404	1	326	gi 1303816	Yqez [Bacillus subtilis]	63	35	728
405	3	2101	gi 1303914	Yqhv [Bacillus subtilis]	63	42	387
406	1	451	gi 112152	sulfate permease (gig start codon) [Synecococcus PCC6301] pir A3010 GYCS7 sulfate transport protein - Synecococcus sp. PCC 7942	63	43	225
415	2	1048	gi 1205402	transport ATP-binding protein [Haemophilus influenzae]	63	41	1671
426	4	3575	gi 193268	28-kilodalton protein [Streptococcus pneumoniae] sp P42362 P29K_STRPM_19 KD MEMBRANE PROTEIN IN PSAA 5' REGION ORF1	63	39	897
505	3	1347	gi 1618999	orf4 [Lactobacillus sake]	63	40	849
507	1	2	gi 46917	comk [Bacillus subtilis, E26, Peptide, 192 aa]	63	35	573
562	2	146	gi 43985	nifS-like gene [Lactobacillus delbrueckii]	63	45	919
675	1	427	gi 1510994	aspartate aminotransferase [Methanococcus jannaschii]	63	29	213
686	1	3	gi 1517356	nitrate reductase (NADH) [Lotus japonicus]	63	52	228
701	1	3	gi 881940	NorO protein [Paracoccus denitrificans]	63	41	390
720	1	2	gi 47168	open reading frame [Streptomyces lividans]	63	35	359
779	1	571	gi 1261932	unknown [Mycobacterium tuberculosis]	63	41	285
907	1	321	gi 149445	ORF1 [Lactococcus lactis]	63	27	300
972	1	794	gi 1511235	M jannaschi: predicted coding region MJ232 [Methanococcus jannaschii]	63	27	386
1085	1	1154	gi 1204277	hypothetical protein [GB U00019.14] [Haemophilus influenzae]	63	38	527
1094	1	3	gi 700943	urea amidolyase [Bacillus subtilis]	63	39	540
1108	1	3	gi 549892 S498	regulation protein - Bacillus subtilis	63	44	440
1113	1	1231	gi 1493017	endocarditis specific antigen [Enterococcus faecalis]	63	45	615

TABLE 2

5. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	maab gene name	% sim	% ident	length (nt)
1100	1	3	695	[sp P33940 YQJH]	HYPOTHETICAL 54.3 KD PROTEIN IN ECO-ALKB INTERGENIC REGION.	63	46	693
1325	1	1	204	[gi 928989]	pl00 protein [Borrelia burgdorferi]	63	30	204
1814	1	3	245	[gi 1303914]	YqjY [Bacillus subtilis]	63	34	243
2021	1	498	250	[pir C33696 C334]	hisc homolog - Bacillus subtilis	63	46	249
2325	1	2	193	[gi 436132]	product is similar to TrpA of transposon Tn556 from Staphylococcus aureus [Clostridium butyricum]	63	40	192
2335	1	1	195	[gi 1184298]	flagellar H5-ring protein [Borrelia burgdorferi]	63	47	195
2406	1	451	227	[gi 1041785]	rhoptry protein [Plasmodium yoelii]	63	33	225
2961	2	136	360	[gi 312463]	carbamoyl-phosphate synthase (glutamine-hydrolysing) [Bacillus aldolyticus]	63	52	225
2965	1	1	402	[gi 1407784]	orf-1; novel antigen [Staphylococcus aureus]	63	50	402
2987	1	583	293	[gi 1224069]	amidase [Moraxella catarrhalis]	63	35	291
2994	1	266	135	[gi 836646]	phosphoribosylformimino-praic ketoisomerase [Rhodococcus phaeoideus]	63	51	132
3043	1	440	252	[gi 1480237]	phenylacetaldehyde dehydrogenase [Escherichia coli]	63	40	169
3078	1	609	400	[gi 1487982]	intrinsic membrane protein [Mycoplasma hominis]	63	36	210
3139	1	2	217	[gi 439126]	glutamate synthase (NADPH) [Azospirillum brasilense] pir[A49916]A49916	63	47	216
3625	1	793	198	[gi 623073]	glutamate synthase (NADPH) [EC 1.4.1.13] - Azospirillum brasilense	63	48	196
3658	1	1	199	[gi 1303697]	Yrka [Bacillus subtilis]	63	37	199
3659	1	3	395	[gi 1256135]	Ybdf [Bacillus subtilis]	63	48	393
3783	1	720	361	[gi 1256903]	Pyruvate decarboxylase isozyme 2 (Swiss Prot. accession number P16467) [Saccharomyces cerevisiae]	63	34	360
3900	1	338	171	[sp P10537 AHYB]	BETA-AMYLASE (EC 3.2.1.2) (1,4-ALPHA-D-GLUCAN MALTOHYDROLASE)	63	54	168
4309	1	3	176	[pir A37967 A379]	neural cell adhesion molecule Ng-CAM precursor - chicken	63	57	174
4367	1	1	195	[gi 1321932]	Per6p gene product [Pichia pastoris]	63	30	195
4442	1	1	312	[gi 151259]	HM-CoA reductase (EC 1.1.1.88) [Pseudomonas nevalonii] pir[A44756]A44756	63	51	312
4468	1	6	308	[gi 296464]	hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) [Pseudomonas sp.]	63	36	303
33	3	1411	2400	[gi 153675]	ATPase [Lactococcus lactis]	62	46	990
36	9	5945	6218	[gi 1490521]	tagatose 6-P kinase [Streptococcus mutans]	62	51	234
				[HMSH]	[Homo sapiens]	62	51	234

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
37	1	2	721	gi 1107531	ceiE gene product (Campylobacter coli)	62	33	720
38	15	10912	11589	gi 1222058	H influenzae predicted coding region HIN1279 (Haemophilus influenzae)	62	38	678
38	25	19526	20322	gi 695280	ORF2 (Alcaligenes eutrophus)	62	41	804
57	2	7523	1780	gi 171234	orf1 (Haemophilus influenzae)	62	55	744
57	9	6646	6350	gi 508174	eltB domain of PTS-dependent Gal transport and phosphorylation Escherichia coli	62	35	297
58	1	2	559	gi 755152	highly hydrophobic integral membrane protein (Bacillus subtilis)	62	34	558
67	10	8230	9014	gi 470683	Shows similarity with ATP-binding proteins from other ABC-transport proteins. Swiss Prot Accession Numbers P24137, P08007, P04285, P24136 Escherichia coli	62	31	765
69	8	8315	7494	gi 46816	actVA 4 gene product (Streptomyces coelicolor)	62	44	822
80	3	1793	1320	gi 39993	UDP-N-acetylmuramoylalanine-D-glutamate ligase (Bacillus subtilis)	62	43	474
87	7	7034	9205	gi 217191	5'-nucleotidase precursor (Vibrio parahaemolyticus)	62	38	2172
100	3	4051	3089	gi 1511047	phosphoglycerate dehydrogenase (Methanococcus jannaschii)	62	42	963
102	1	2	520	gi 151655	mismatch repair protein (Streptococcus pneumoniae)	62	34	519
112	2	466	1068	gi 153741	ATP-binding protein (Streptococcus mutans)	62	37	603
114	7	6855	7562	gi 1204866	L-fucose operon activator (Haemophilus influenzae)	62	38	708
116	4	6823	5633	gi 677947	AppC (Bacillus subtilis)	62	37	1191
124	8	6855	6004	gi 853777	product similar to E.coli PRA2 protein (Bacillus subtilis)	62	44	852
148	1	24	554	gi 467436	unknown (Bacillus subtilis)	62	30	531
149	20	7591	6725	gi 1205807	replicative DNA helicase (Haemophilus influenzae)	62	41	867
163	3	1503	1153	gi 40067	X gene product (Bacillus sphaericus)	62	42	331
164	15	14673	15632	gi 42219	P15 gene product (AA 1 - 314) (Escherichia coli)	62	38	960
165	2	1166	1447	gi 403936	phenylalanyl-tRNA synthetase alpha subunit (Gly294 variant) unidentified (cloning vector)	62	38	282
166	2	2084	5089	gi 308861	CTG start codon (Lactococcus lactis)	62	43	3086
171	1	1225	614	gi 1046053	hypothetical protein (sp.P12049) (Mycolasma genitalium)	62	41	612

TABLE 2

5. aureus - putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
183	1	2521	1310	gi1143045	hemY [Bacillus subtilis]	62	45	1212
200	1	1	956	gi1142439	ATP-dependent nuclease [Bacillus subtilis]	62	32	954
237	2	935	1966	gi141495	hisC protein [Escherichia coli]	62	44	1032
261	3	4008	2605	gi1143121	ORF A: putative [Bacillus firmus]	62	42	1404
299	8	4477	4719	gi1467441	expressed at the end of exponential growth under conditions in which he enzymes of the TCA cycle are repressed [Bacillus subtilis] gi1467441	62	47	233
					expressed at the end of exponential growth under conditions in which the enzymes of the TCA cycle are repressed [Bacillus subtilis]			
364	6	5018	1819	gi1152015	FemA protein [Staphylococcus aureus]	62	43	1200
324	1	2	262	gi1142717	cytochrome aa3 controlling protein [Bacillus subtilis] p1rA13960[A13960 cta protein - Bacillus subtilis sp[P12946[CTAA_BACSU CYTOCHROME AA3 CONTROLLING PROTEIN	62	33	261
325	2	269	1207	gi1581088	methionyl-tRNA formyltransferase [Escherichia coli]	62	39	939
332	6	4694	4631	gi1149980	uridine 5'-monophosphate synthase [Methanococcus jannaschii]	62	36	264
355	1	2	370	gi1145925	fecB [Escherichia coli]	62	32	369
365	8	6628	6804	gi1413943	ipa-19d gene product [Bacillus subtilis]	62	54	177
369	2	2744	1626	p1r[A13577]A435	regulatory protein pfor - Clostridium perfringens	62	42	1119
370	1	34	264	gi140665	beta-glucosidase [Clostridium thermocellum]	62	37	311
415	3	2109	3176	gi11205401	transport ATP-binding protein [Haemophilus influenzae]	62	35	468
429	1	1578	790	gi11046024	Na+ ATPase subunit J [Mycoplasma genitalium]	62	40	749
444	2	704	1369	gi1581510	modulation gene; integral membrane protein; homology to Rhizobium eguminosarum nodi [Rhizobium loti]	62	37	656
477	2	751	1869	p1r[A18440]A484	ring-infected erythrocyte surface antigen 2, RESA-2 - Plasmodium falciparum	62	44	1119
485	1	241	1707	gi117934	betaine aldehyde dehydrogenase [Beta vulgaris]	62	43	1457
487	3	1141	1311	gi1149445	ORF1 [Lactococcus lactis]	62	31	171
494	2	1134	1313	gi1166835	ribulose biphosphate carboxylase/oxygenase activase [Arabidopsis thaliana]	62	37	180
518	1	193	882	gi1153491	O-methyltransferase [Streptomyces glaucescens]	62	39	690
534	2	369	2522	gi11480429	putative transcriptional regulator [Bacillus stearothermophilus]	62	35	2154
551	6	4371	4820	gi1211113	ferric uptake regulation protein [Campylobacter jejuni]	62	37	450
574	1	1	570	gi1153600	enterotoxin B [Staphylococcus aureus]	62	43	570

TABLE 2

5 aureus : putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
590	2	344	1171	gi 40367	ORF1 (Clostridium acetobutylicum)	62	37	828
655	1	396	830	gi 147195	phnB protein (Escherichia coli)	62	44	435
656	1	2	478	gi 1205451	cell division inhibitor (Haemophilus influenzae)	62	36	477
676	1	692	348	gi 1511613	methyl coenzyme M reductase system, component A2 (Methanococcus jannaschii)	62	36	365
687	1	493	248	gi 49272	Asparaginase (Bacillus licheniformis)	62	48	246
700	2	267	944	gi 1205822	hypothetical protein (CB-X75627_4) (Haemophilus influenzae)	62	40	678
840	2	1715	1041	gi 1045865	M. genitalium predicted coding region NG14 (Mycoplasma genitalium)	62	36	675
866	4	898	1491	gi 1144332	deoxyuridine nucleotidylhydrolase (Homo sapiens)	62	38	594
916	1	35	400	gi 413931	ipa-7d gene product (Bacillus subtilis)	62	45	366
1071	1	1	771	gi 1510649	aspartokinase 1 (Methanococcus jannaschii)	62	40	771
1084	1	19	609	gi 488011	AgoX-1 antigen (human, infertile patient, testis, peptide, 505 aa)	62	39	591
1103	1	3	203	gi 581261	ORF homologous to E. coli metB (Herpetosiphon aurantiacus) pIR S14030 S14030 (hypothetical protein - Herpetosiphon aurantiacus fragment)	62	51	201
1217	1	463	233	gi 460025	ORF2, putative (Streptococcus pneumoniae)	62	41	211
1533	1	644	414	gi 413968	ipa-4d gene product (Bacillus subtilis)	62	48	211
1537	1	3	257	gi 1510641	alanine-tRNA synthetase (Methanococcus jannaschii)	62	29	255
2287	1	3	161	gi 485956	ImjC gene product (Mycobacterium tuberculosis)	62	45	159
2386	1	3	245	gi 285708	non-toxic component (Clostridium botulinum)	62	31	243
2484	1	331	167	gi 1142092	DNA-repair protein (recA) (Anabaena variabilis)	62	35	165
2490	1	798	400	gi 581648	lepB gene product (Staphylococcus epidermidis)	62	42	399
3016	1	596	300	gi 710022	uroporphyrinogen III (Bacillus subtilis)	62	51	297
3116	1	1	213	gi 466883	InfS: B1496_C2_193 (Mycobacterium leprae)	62	44	213
3297	1	823	413	gi 475715	acetyl coenzyme A acetyltransferase (thiolase) (Clostridium acetobutylicum)	62	42	411
3609	1	31	276	gi 1140830	homologous to H-acyl-L-amino acid amidohydrolase of Bacillus stearothermophilus (Bacillus subtilis)	62	48	246
3665	2	584	402	gi 151259	HMG-CoA reductase (EC 1.1.1.88) (Pseudomonas mevaloni) pIR A44756 A44756 (hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.)	62	40	183
3733	1	3	374	gi 1353197	thioredoxin reductase (Eubacterium acidaminophilum)	62	42	372

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
3898	1	1	237	gi153675	lagatase 6-p kinase [Streptococcus mutans]	62	45	237
4027	1	283	143	gi1330705	homologue to gene 30 (aa 1-59), putative [Bovine herpesvirus 1]	62	43	141
4109	1	727	365	gi141748	hsm protein (AA 1-520) [Escherichia coli]	62	45	363
4303	1	1	303	gi1303813	Yqew [Bacillus subtilis]	62	43	303
4380	1	530	267	gi1235684	nevalonate pyrophosphate deacetylase [Saccharomyces cerevisiae]	62	55	264
4494	1	2	256	gi1510892	enterotoxin H [Staphylococcus aureus]	62	34	255
4598	1	411	223	gi1763513	ORF4, putative [Streptomyces violaceoruber]	62	45	189
4624	1	1	222	gi141748	hsm protein (AA 1-520) [Escherichia coli]	62	45	222
5	5	4288	3912	gi1928831	ORF95, putative [Lactococcus lactis phage BK5-T]	61	36	357
11	1	320	162	gi133356[C33]	prothymosin alpha homolog (clone 32) - human (fragment)	61	33	159
16	11	10991	11938	gi1205391	hypothetical protein (SP-P31995) [Haemophilus influenzae]	61	44	948
32	1	283	801	gi1066504	exo-beta 1,3 glucanase [Orchilobolus caribonum]	61	50	519
38	3	616	1107	gi1510864	glutamine transport ATP-binding protein Q [Methanococcus jannaschii]	61	41	492
45	4	3082	4038	gi1105686	Prox [Bacillus subtilis]	61	45	957
48	8	7118	7504	gi1498839	ORF2 [Clostridium perfringens]	61	33	387
51	9	4605	5570	gi1388269	trac [Plasmodium falciparum]	61	42	966
60	6	1889	2243	gi1205893	hypothetical protein (GB-U00011_3) [Haemophilus influenzae]	61	32	555
62	9	5559	5122	gi1854656	Na/H antiporter system OMF2 [Bacillus alcalophilus]	61	38	438
67	5	4330	5646	gi1466612	nika [Escherichia coli]	61	36	1317
74	2	2400	1504	gi1204846	carbamate kinase [Haemophilus influenzae]	61	40	857
85	1	2198	1101	gi1498756	amidophosphoribosyltransferase PurF [Rhizobium etli]	61	41	1098
86	4	1995	1582	gi1499931	M. jannaschii predicted coding region M1083 [Methanococcus jannaschii]	61	44	414
97	1	74	649	gi1518679	orf [Bacillus subtilis]	61	44	576
99	2	2454	1990	gi1413958	ipa-34d gene product [Bacillus subtilis]	61	18	465
124	7	6223	5123	gi1556681	Similar to Saccharomyces cerevisiae SUA5 protein [Bacillus subtilis] gi1549158[S49158] ipc-29d protein - Bacillus subtilis ap19153[YMUC_BACSU HYPOTHETICAL 37.0 KD PROTEIN IN SP01R-GLYC NTERGENIC REGION	61	46	1101
125	4	1668	2531	gi1491643	ORF4 gene product [Chloroflexus aurantiacus]	61	43	864

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	A sim	A ident	length (nt)
112	1	1250	627	pir P0259 P002	hypothetical protein 1 - Enterococcus faecalis plasmid pAM-beta-1 (fragment)	61	43	624
119	9	3617	3075	gi 1144132	deoxyuridine nucleotidohydrolase [Homo sapiens]	61	40	543
149	12	8690	7869	gi 160047	p101/acidic basic repeat antigen [Plasmodium falciparum] pir A29232 A29232	61	35	822
168	3	1915	2361	gi 11499694	101K malaria antigen precursor - Plasmodium elciaparum (strain Camp)	61	41	447
171	9	9675	7948	gi 467416	HIT protein, member of the HIT-family [Methanococcus jannaschii]	61	38	1228
174	3	1042	2140	gi 216374	similar to SpoVB [Bacillus subtilis]	61	49	1209
190	4	5034	4111	gi 409286	glucyl 7-ACA acylase precursor [Bacillus laterosporus]	61	37	924
216	1	2	190	gi 415861	BerU [Bacillus subtilis]	61	29	183
227	7	4161	5048	gi 216341	eukaryotic initiation factor 2 beta (eIF-2 beta) [Cryptosporidium parvum]	61	41	888
238	4	1959	3047	gi 409543	ORF for methionine amino peptidase [Bacillus subtilis]	61	38	1089
247	1	2	694	gi 537231	ChRC protein [Erwinia chrysanthemi]	61	38	693
247	2	678	1034	gi 142226	ORF_1579 [Escherichia coli]	61	40	357
257	2	3523	2627	gi 699379	ChVD protein [Agrobacterium tumefaciens]	61	40	897
268	2	3419	3051	gi 40364	glvz-1 protein [Mycobacterium leprae]	61	41	369
275	4	4621	4827	gi 1204848	ORFA1 [Clostridium acetobutylicum]	61	36	207
277	1	1	1845	gi 794897	hypothetical protein (GP-H87049.57) [Haemophilus influenzae]	61	45	1845
278	9	8001	7032	gi 467462	beta-N-acetylhexosaminidase [Streptococcus pneumoniae] pir A56390 A56390	61	41	972
278	10	9878	8535	gi 1205919	mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - Streptococcus pneumoniae	61	38	1344
283	1	1	366	gi 755607	cysteine synthetase A [Bacillus subtilis]	61	36	366
288	2	1918	1496	gi 388108	Na <sup>+</sup> and Cl <sup>-</sup> dependent gamma-aminobutyric acid transporter [Haemophilus influenzae]	61	43	423
291	1	86	334	gi 454265	cell wall enzyme - Enterococcus faecalis	61	38	249
318	1	1104	694	gi 290531	PBP3 [Pectunia hybrida]	61	47	411
330	2	1912	1190	gi 1001805	similar to beta-glucoside transport protein [Escherichia coli] sp P31451 PTIB-ECOLI PTS SYSTEM, ARABUTIN-LIKE IIB COMPONENT PHOSPHOTRANSFERASE ENZYME II, B COMPONENT (EC 2.7.1.69)	61	41	723

TABLE 2



S. aureus - putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
385	2	1513	1025	gi 531098	DnaD protein [Bacillus subtilis]	61	42	489
426	3	794	399	gi 1303853	YggF [Bacillus subtilis]	61	44	396
438	3	810	1421	gi 1293680	AbaA2 [Streptomyces coelicolor]	61	36	612
454	1	1580	792	gi 733522	phosphatidylinositol-4,5-diphosphate 3-kinase [Dictyostellium discoideum]	61	30	789
464	2	784	560	gi 1123120	CS187.5 gene product [Caenorhabditis elegans]	61	38	225
470	8	6077	7357	gi 623073	ORF160, putative [Bacteriophage LU-H]	61	47	1281
509	1	554	279	gi 467484	unknown [Bacillus subtilis]	61	45	276
555	3	1916	1296	gi 141800	anthranilate synthase glutamine amidotransferase [Acinetobacter calcoaceticus]	61	42	621
569	1	1711	857	gi 467090	R2315_C2_195 [Mycobacterium leprae]	61	47	855
585	2	961	803	sp P36686 SURE_	SURVIVAL PROTEIN SURE HOMOLOG (FRAGMENT)	61	33	159
592	3	1694	1422	gi 1221602	immunity repressor protein [Haemophilus influenzae]	61	32	273
603	1	43	357	gi 507738	Ilp [Vibrio parahaemolyticus]	61	33	335
669	1	2467	1235	gi 1146243	22.4% identity with Escherichia coli DNA-damage inducible protein ... putative [Bacillus subtilis]	61	37	1233
675	3	805	1101	gi 403373	glycerophosphoryl diester phosphodiesterase [Bacillus subtilis] pfr[S37251 S3725] glycerophosphoryl diester phosphodiesterase - acillus subtilis	61	36	297
703	1	1656	829	gi 537181	ORF_1470 [Escherichia coli]	61	32	828
728	1	1628	816	gi 806281	DNA polymerase I [Bacillus stearothermophilus]	61	39	813
821	1	61	318	gi 709992	hypothetical protein [Bacillus subtilis]	61	38	258
856	2	2313	1567	gi 609310	portal protein gp1 [Bacteriophage HK97]	61	40	147
923	1	1081	542	gi 143213	putative [Bacillus subtilis]	61	38	540
1124	1	59	370	gi 1107541	C3309.8 [Caenorhabditis elegans]	61	26	312
1492	1	548	276	gi 406397	unknown [Mycoplasma genitalium]	61	32	273
1602	1	46	318	gi 733522	phosphatidylinositol-4,5-diphosphate 3-kinase [Dictyostellium discoideum]	61	34	273
2500	1	577	290	gi 1045964	hypothetical protein (U8:U14003.297) [Mycoplasma genitalium]	61	31	288
2968	1	2	808	gi 397526	clumping factor [Staphylococcus aureus]	61	55	807
3076	1	3	248	gi 169373	ORF 1 [Lactococcus lactis]	61	41	246

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
3609	2	207	gi 1408501	homologous to N-acetyl-L-amino acid amidohydrolase of <i>Bacillus stearothermophilus</i> [Bacillus subtilis]	61	39	175
3662	1	1477	gi 1303813	Yqew [Bacillus subtilis]	61	42	738
3672	1	2	gi 784897	beta-N-acetylhexosaminidase (Streptococcus pneumoniae) pir[A56390/A56390] mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - streptococcus pneumoniae	61	50	441
3724	1	2	gi 1009366	Respiratory nitrate reductase [Bacillus subtilis]	61	41	219
3728	1	398	gi 677943	AppD [Bacillus subtilis]	61	46	396
3884	1	3	gi 784897	beta-N-acetylhexosaminidase (Streptococcus pneumoniae) pir[A56390/A56390] mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - streptococcus pneumoniae	61	47	399
3971	1	383	gi 784897	beta-N-acetylhexosaminidase (Streptococcus pneumoniae) pir[A56390/A56390] mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - streptococcus pneumoniae	61	45	381
4038	1	661	gi 1339970	large subunit of NAM-dependent glutamate synthase (Plectonema boryanum)	61	24	303
4041	1	546	gi 413953	lin-29d gene product [Bacillus subtilis]	61	48	273
4047	1	402	gi 528991	unknown [Bacillus subtilis]	61	42	402
4102	1	345	gi 976025	HrSA [Escherichia coli]	61	46	345
4155	1	336	gi 784897	beta-N-acetylhexosaminidase (Streptococcus pneumoniae) pir[A56390/A56390] mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - streptococcus pneumoniae	61	50	336
4288	1	463	gi 450688	hadh gene of Ecoptri gene product [Escherichia coli] pir[S38437/S38437] hadH protein - Escherichia coli pir[S09629/S09629] hypothetical protein A - Escherichia coli (S09 40-520)	61	38	231
4374	1	542	gi 784897	beta-N-acetylhexosaminidase (Streptococcus pneumoniae) pir[A56390/A56390] mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - streptococcus pneumoniae	61	50	270
4389	1	172	gi 147516	ribokinase [Escherichia coli]	61	35	171
4621	1	268	gi 784897	beta-N-acetylhexosaminidase (Streptococcus pneumoniae) pir[A56390/A56390] mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - streptococcus pneumoniae	61	47	267
4663	1	227	gi 976025	HrSA [Escherichia coli]	61	50	201
4	6	6663	gi 1408501	homologous to N-acetyl-L-amino acid amidohydrolase of <i>Bacillus stearothermophilus</i> [Bacillus subtilis]	60	43	1128

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
11	6	3426	3725	gi 410748	ring-infected erythrocyte surface antigen (Plasmodium falciparum) pif[A25526]A25526 ring-infected erythrocyte surface antigen recursor - Plasmodium falciparum (strain FC27/Papua New Guinea) sp P13810 RESA_PLAFF	60	24	300
					RING-INFECTED ERYTHROCYTE SURFACE ANTIGEN RE			
11	14	11035	10133	gi 1217651	carboxyl reductase (NADPH) (Rattus norvegicus)	60	28	723
16	12	11917	12930	gi 1001453	hypothetical protein [Synchocystis sp.]	60	37	1014
33	1	26	469	gi 388109	regulatory protein [Enterococcus faecalis]	60	41	444
37	13	10914	9834	gi 1316056	Orf1 [Bacillus subtilis]	60	40	981
39	4	4364	4522	gi 4872	ORF 4 [Saccharomyces kluyveri]	60	47	159
41	1	2047	1025	gi 142822	D-alanine racemase cds [Bacillus subtilis]	60	39	1023
43	4	2474	3607	gi 468046	para-nitrobenzyl esterase [Bacillus subtilis]	60	40	1134
44	10	6755	7769	gi 414234	thiF [Escherichia coli]	60	52	1014
45	10	8874	9074	gi 343949	Var1(40 D) [Saccharomyces cerevisiae]	60	44	201
56	18	27842	26430	gi 468764	mncR gene product [Rhizobium meliloti]	60	35	1413
60	2	173	388	gi 1303864	YggQ [Bacillus subtilis]	60	33	216
63	2	357	1619	gi 467124	ureD_B229-C1_234 [Mycobacterium lepreae]	60	43	1263
69	1	787	395	gi 1518653	OafA [Salmonella typhimurium]	60	36	393
88	1	1	1188	gi 480429	putative transcriptional regulator [Bacillus stearothermophilus]	60	30	1188
92	6	4735	3881	gi 349227	transmembrane protein [Escherichia coli]	60	37	855
92	7	5996	4923	gi 466613	nlkB [Escherichia coli]	60	38	1074
93	1	949	476	gi 1510925	coenzyme F420-reducing hydrogenase, beta subunit [Methanococcus jannaschii]	60	27	474
96	6	7166	7578	gi 972715	acetylase protein [Carnobacterium plautii]	60	30	213
98	6	3212	4069	gi 467425	unknown [Bacillus subtilis]	60	42	858
102	10	7158	7430	gi 143092	acetolactate synthase small subunit [Bacillus subtilis]	60	37	273
					sp P37252 ILUN_BACSU ACETOLACTATE SYNTHASE SMALL SUBUNIT (EC 1.3.18) [ILUN] (ACETOHYDROXY-ACID SYNTHASE SMALL SUBUNIT) (ALS)			
109	11	9127	10515	gi 1255259	o-succinylbenzoic acid (OSB) CoA ligase [Staphylococcus aureus]	60	28	1189
109	12	10499	11656	gi 141954	beta-ketothiolase [Alcaligenes eutrophus]	60	41	1158
119	2	4630	3134	gi 1524280	unknown [Mycobacterium tuberculosis]	60	45	1497

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
121	9	6957	7646	[gi1107529]	ceuC gene product [Campylobacter coli]	60	35	690
140	7	7704	6013	[gi146547]	kdpA [Escherichia coli]	60	45	1692
145	1	2	701	[gi1460077]	unknown [Mycobacterium tuberculosis]	60	23	702
150	3	2809	2216	[gi1146230]	putative [Bacillus subtilis]	60	40	594
157	2	1389	961	[gi1103975]	YqjX [Bacillus subtilis]	60	30	479
158	5	5125	4769	[gi1449288]	unknown [Mycobacterium tuberculosis]	60	36	357
159	1	511	257	[gi1580932]	murD gene product [Bacillus subtilis]	60	43	255
160	1	159	1187	[gi1204532]	hypothetical protein (GB:U19201.29) [Haemophilus influenzae]	60	34	1029
161	14	8249	7866	[gi1496003]	ORF3, PepY, putative oligoendopeptidase based on homology with Lactococcus lactis PepF (GenBank Accession Number 232522) [Caldicellulosiruptor saccharolyticus]	60	34	384
172	3	1331	2110	[gi1485280]	128.2 kDa protein (Streptococcus pneumoniae)	60	33	780
173	2	4082	2460	[gi1524397]	glycine betaine transporter Opud [Bacillus subtilis]	60	41	1623
173	3	5163	4953	[gi1100737]	NAIP dependent leukotriene B <sub>4</sub> 12-hydroxylase [Sus scrofa]	60	44	1011
198	1	3	995	[gi1413941]	lpa-19d gene product [Bacillus subtilis]	60	42	993
201	4	3641	4573	[sp137028]YADT_	HYPOTHETICAL 29.4 KD PROTEIN IN HEMU-PFS INTERGENIC REGION PRECURSOR	60	37	933
201	3	3269	2415	[gi1927798]	109719.34p, CAI: 0.14 [Saccharomyces cerevisiae]	60	43	875
206	9	12234	12515	[sp137347]YBCD_	HYPOTHETICAL 21.8 KD PROTEIN IN ASP5 5'-REGION	60	47	282
212	4	1213	1410	[gi1332711]	hemagglutinin-neuraminidase fusion protein [Human parainfluenza virus 3]	60	34	198
214	1	65	1153	[gi1204366]	hypothetical protein (GB:U14003.130) [Haemophilus influenzae]	60	36	1089
237	1	2	937	[gi1149377]	HLSD [Lactococcus lactis]	60	40	936
241	6	5696	4998	[gi1046160]	hypothetical protein (GB:U00021.5) [Mycoplasma genitalium]	60	37	699
260	6	5919	6485	[gi1431950]	similar to a B-subtilis gene (GB: BACHEMIV.5) [Clostridium acetabutylicum]	60	35	567
264	1	2432	1218	[gi1397526]	clumping factor [Staphylococcus aureus]	60	53	1215
267	1	3	1409	[gi1148316]	NaN-antiporter protein [Enterococcus hirae]	60	27	1407
275	3	3804	4595	[pir-F36889]F368	leuD 3'-region hypothetical protein - Lactococcus lactis subsp. lactis (strain IL1403)	60	35	792
291	3	860	1198	[gi1208889]	coded for by C. elegans cDNA yk130a2.5; contains C2H2-type zinc fingers [Caenorhabditis elegans]	60	33	339

TABLE 2

S. aureus - putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
307	6	3421	3176	gi1070014	protein-dependent [Bacillus subtilis]	60	36	246
316	8	4957	5823	gi1413952	lipa-28d gene product [Bacillus subtilis]	60	41	867
328	4	2996	3484	gi1120484	membrane-associated component, branched amino acid transport system [Haemophilus influenzae]	60	39	489
332	5	4887	4363	gi11205449	colicin V production protein (pur regulon) [Haemophilus influenzae]	60	37	525
337	1	1062	532	gi11487842	single-stranded DNA-specific exonuclease [Escherichia coli]	60	41	531
375	2	96	362	gi114057	adenylyl cyclase gene product [Saccharomyces kluyveri] t1JQ114510BYK adenylyl cyclase [EC 4.6.1.1] - yeast ccharomyces kluyveri	60	47	267
397	1	66	416	gi1709999	glucuronate dehydratase [Bacillus subtilis]	60	37	351
409	1	2	163	gi1499700	glycogen phosphorylase [Saccharomyces cerevisiae]	60	35	162
433	4	914	1237	gi11196899	unknown protein [Staphylococcus aureus]	60	36	324
433	7	3838	3620	sp1P1222 YCP1_	HYPOTHETICAL 226 KD PROTEIN (ORF 1901)	60	31	219
470	2	622	945	pir1S10782 S107	integrin homolog - yeast [Saccharomyces cerevisiae]	60	31	324
500	1	118	606	gi1467407	unknown [Bacillus subtilis]	60	36	489
503	3	752	982	gi1167835	myosin heavy chain [Dictyostellium discoideum]	60	34	231
505	4	2238	3563	gi11510732	NADH oxidase [Methanococcus jannaschii]	60	26	1326
523	1	3	1043	gi1143331	alkaline phosphatase regulatory protein [Bacillus subtilis] pir1A27650 A27650 regulatory protein phoB - Bacillus subtilis sp1P23515 PHOR_BAC50 ALKALINE PHOSPHATASE SYNTHESIS SENSOR PROTEIN HOR IEC 2.7.3.-1	60	41	1041
543	1	1	465	gi11511103	cobalt transport ATP-binding protein O [Methanococcus jannaschii]	60	40	465
545	1	1	726	gi11498192	putative [Pseudomonas aeruginosa]	60	40	726
556	1	2	1054	gi11477402	txe gene product [Bordetella pertussis]	60	42	1053
576	1	974	489	gi11205129	ph. influenzae predicted coding region H10882 [Haemophilus influenzae]	60	42	486
594	1	1	624	gi11212755	adenylyl cyclase [Aeromonas hydrophila]	60	35	624
604	1	3	530	gi1145925	fecB [Escherichia coli]	60	42	528
620	1	926	465	gi11205483	bicyclicomycin resistance protein [Haemophilus influenzae]	60	33	462
630	2	871	1122	gi11486242	unknown [Bacillus subtilis]	60	41	252
645	2	574	425	gi11205136	serine hydroxymethyltransferase [serine methylase] [Haemophilus influenzae]	60	28	150

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match	match gene name	% sim	% ident	length (nt)
684	1	1082	843	gi 1205538		hypothetical protein [GB.U14003.302] [Haemophilus influenzae]	60	39	240
786	1	967	485	gi 1402944		orf8M gene product [Bacillus subtilis]	60	46	483
844	1	586	346	gi 790943		urea amidolyase [Bacillus subtilis]	60	40	243
851	1	1	726	gi 159661		GMP reductase [Ascaris lumbricoides]	60	41	726
971	1	1746	874	gi 1001493		hypothetical protein [Synecocystis sp.]	60	39	873
996	1	1558	839	gi 604926		NADH dehydrogenase, subunit 5 [Schizosaccharomyces pombe] sp P50368 NUSM_SCHCO NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5 [EC 6.5.3.1]	60	39	720
908	2	448	753	gi 662880		novel hemolytic factor [Bacillus cereus]	60	31	306
979	1	2	595	gi 1429255		putative, orf1 [Bacillus subtilis]	60	30	594
1078	1	649	502	gi 581055		inner membrane copper tolerance protein [Escherichia coli] gi 871029 disulphide isomerase like protein [Escherichia coli] pir S47295 S47295 inner membrane copper tolerance protein - acherichia coli	60	40	168
1112	1	1150	620	gi 407885		ORF3 [Streptomyces griseus]	60	34	531
1115	1	484	275	gi 1171407		VpsM [Saccharomyces cerevisiae]	60	36	210
1146	1	17	562	gi 1239981		hypothetical protein [Bacillus subtilis]	60	36	546
1291	1	716	360	pir S57530 S575		carboxyl esterase - Acinetobacter calcoaceticus	60	30	357
1332	1	336	169	gi 1222056		aminotransferase [Haemophilus influenzae]	60	44	168
1429	1	3	146	gi 1205619		ferritin like protein [Haemophilus influenzae]	60	39	144
1722	1	570	286	gi 240052		dihydroflavonol-4-reductase, DFR [Hordeum vulgare-barley, cv. Gula, eptide, 354 aa]	60	36	285
2350	1	385	200	gi 497626		ORF 1 [Plasmid pAQ1]	60	20	186
2936	1	519	310	gi 508981		prephenate dehydratase [Bacillus subtilis]	60	48	210
3027	1	568	302	gi 146199		putative [Bacillus subtilis]	60	37	267
3084	1	20	208	gi 1407784		orf-1, novel antigen [Staphylococcus aureus]	60	51	189
3155	1	2	226	gi 1046097		cytadherence-accessory protein [Mycoplasma genitalium]	60	34	225
3603	1	368	186	gi 510108		mitochondrial long-chain enoyl-CoA hydratase/3-hydroxycyl-CoA enoylhydrolase alpha-subunit [Rattus norvegicus]	60	42	183
3665	1	486	244	gi 151259		HMG-CoA reductase [EC 1.1.1.88] [Pseudomonas mavalonil] pir A44756 A44756 hydroxymethylglutaryl-CoA reductase [EC 1.1.1.88] Pseudomonas sp.	60	42	243
3747	1	3	146	gi 474192		lucC gene product [Escherichia coli]	60	36	144

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match position	match gene name	% sim	% ident	length (nt)
1912	1	3	335	gi 488695	novel antigen, orf-2 [Staphylococcus aureus]	60	44	333
4072	1	3	272	gi 405829	yleH [Escherichia coli]	60	33	270
4134	1	510	352	gi 780656	chemoreceptor protein [Rhizobium leguminosarum bv. viciae] gi 780656 chemoreceptor protein [Rhizobium leguminosarum bv. lotiae]	60	28	159
4207	2	677	402	gi 602031	similar to trimethylamine DH [Mycoplasma capricolum] pir 549950 549950 (SGC3) (fragment)	60	41	276
4243	1	127	324	gi 899317	peptide synthetase module [Microcystis aeruginosa] pir 549111 549111 probable amino acid activating domain - microcystis aeruginosa (fragment) (SUB 144-528)	60	42	198
4710	1	624	313	gi 308980	pheB [Bacillus subtilis]	60	28	312
4745	1	343	173	gi 510108	mitochondrial long-chain enoyl-CoA hydratase/3-hydroxycyl-CoA dehydrogenase alpha-subunit [Rattus norvegicus]	60	42	171
4782	1	498	280	gi 47382	acyl-CoA-dehydrogenase [Streptococcus purpurascens]	60	48	219
4474	1	53	223	gi 510108	mitochondrial long-chain enoyl-CoA hydratase/3-hydroxycyl-CoA dehydrogenase alpha-subunit [Rattus norvegicus]	60	42	171
23	4	4518	3523	gi 426446	VipB protein [Salmonella typhi]	59	39	996
33	2	707	1483	gi 548604 5486	hypothetical protein - Mycoplasma capricolum (SGC3) (fragment)	59	33	777
33	5	4651	5853	gi 6721	P5982.3 [Caenorhabditis elegans]	59	33	1203
37	2	3228	2299	gi 142813	ORF2 [Bacillus subtilis]	59	37	930
38	21	16784	16593	gi 912576	BIP [Phaeodactylum tricornutum]	59	40	192
52	3	2648	2349	gi 536972	ORF_090a [Escherichia coli]	59	44	300
54	12	14181	13402	gi 483940	transcription regulator [Bacillus subtilis]	59	37	780
57	3	4397	3339	gi 508176	Gat-1-P-DH, NAD dependent [Escherichia coli]	59	40	1059
66	1	986	495	gi 1303901	Yqht [Bacillus subtilis]	59	34	492
67	7	6552	7460	gi 912461	nLKC [Escherichia coli]	59	37	909
70	7	5383	6366	gi 1399822	PhoD precursor [Rhizobium meliloti]	59	46	984
78	1	1	1449	gi 971345	unknown, similar to E.coli cardiolipin synthase [Bacillus subtilis] sp P45860 YME_BACSU HYPOTHETICAL 58.2 PROTEIN IN NARI-ACDA NTERGENIC REGION	59	39	1449
82	10	14329	15534	gi 490328	LORF P (unidentified)	59	44	1206

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
89	2	1602	958	gi1642801	unknown [Saccharomyces cerevisiae]	59	32	645
96	4	4940	5473	gi11333802	protein of unknown function [Rhodobacter capsulatus]	59	33	534
98	1	2	820	gi1467421	similar to B. subtilis DnaH [Bacillus subtilis]	59	34	819
119	1	166	1557	gi1143122	ORF B; putative [Bacillus firmus]	59	36	1392
120	10	6214	6756	gi115354	ORF 55.9 [Bacteriophage 74]	59	39	541
120	16	12476	13510	gi11086575	BotA [Rhizobium meliloti]	59	44	1035
123	1	386	195	gi1984737	catalase [Campylobacter jejuni]	59	38	192
130	1	370	645	gi11256634	25.8% identity over 120 aa with the Synenococcus sp. MpeV protein; putative [Bacillus subtilis]	59	31	276
131	4	5278	5712	gi11510655	hypothetical protein (SP-P42297) [Methanococcus jannaschii]	59	39	435
164	1	3	509	gi11001342	hypothetical protein [Synecocystis sp.]	59	41	507
164	4	1529	2821	gi11205165	hypothetical protein (SP-P37764) [Haemophilus influenzae]	59	35	1293
164	19	19643	21376	gi11001381	hypothetical protein [Synecocystis sp.]	59	34	1734
173	3	4727	3717	gi11184121	auxin-induced protein [Vigna radiata]	59	50	1011
179	2	2218	1688	gi1143036	unidentified gene product [Bacillus subtilis]	59	33	531
195	12	12669	11503	gi1762778	Nifs gene product [Anabaena azollae]	59	41	1167
201	5	4702	5670	gi11510240	hemolysin [Methanococcus jannaschii]	59	32	969
201	7	5719	6315	gi11511456	M. jannaschii predicted coding region MJ1437 [Methanococcus jannaschii]	59	34	597
209	1	102	461	gi11204666	hypothetical protein (CM-X71124_53) [Haemophilus influenzae]	59	42	360
214	3	1050	2234	gi11551531	2-nitropropane dioxygenase [Williopsis saturnus]	59	36	1185
214	5	3393	4135	gi11301709	YrkJ [Bacillus subtilis]	59	32	843
217	2	3381	2167	gi1200489	dcp (CG Site No. 18430) [Escherichia coli]	59	44	1215
237	5	3078	3785	gi1149382	IlvA [Lactococcus lactis]	59	38	708
251	2	376	960	gi11303791	YqoJ [Bacillus subtilis]	59	34	585
286	1	1821	812	gi1146551	transmembrane protein (kdp) [Escherichia coli]	59	31	810
316	5	4978	3860	gi1405879	YehH [Escherichia coli]	59	32	1119
370	3	600	761	gi11303794	Yqem [Bacillus subtilis]	59	35	162

TABLE 2



5. aureus - putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
382	1	1009	506	gi1147513	[orf1] Haemophilus influenzae	59	34	504
391	3	1620	1273	gi1152901	[ORF 3] Spirochaeta aurantia	59	37	348
406	3	2805	1705	gi1709992	hypothetical protein [Bacillus subtilis]	59	34	1101
426	5	3802	3245	gi11204610	[ironIII] dicitrate transport ATP-binding protein PECE Haemophilus influenzae	59	36	558
429	2	1513	1148	gi11064809	homologous to sp-HTRA_ECOLI [Bacillus subtilis]	59	42	366
460	2	708	1301	gi1466882	[ppa1: B1496_C2_189] Mycobacterium leprae	59	37	594
461	4	2212	3135	gi11498295	homoserine kinase homolog [Streptococcus pneumoniae]	59	37	924
473	1	2929	1607	gi1147989	trigger factor [Escherichia coli]	59	40	1323
480	8	5862	6110	gi11205311	[13a]-hydroxymyristol acyl carrier protein dehydrase [Haemophilus influenzae]	59	40	249
521	1	14	1354	pir1A25620A256	[staphylococcalase - Staphylococcus aureus (fragment)]	59	32	1341
534	4	2994	4073	gi1153746	[mannitol-phosphate dehydrogenase [Streptococcus mutans] pir[C44798]C44798 mannitol-phosphate dehydrogenase M1D - streptococcus mutans	59	36	1080
535	1	1	954	gi11469939	[group B oligopeptidase Pept [Streptococcus agalactiae]	59	33	954
551	3	2826	3186	gi11204511	[bacterioferitin comigratory protein [Haemophilus influenzae]	59	45	351
573	2	449	940	gi1386681	[ORF VAL022] [Saccharomyces cerevisiae]	59	36	492
650	1	5	748	gi1396400	[similar to eukaryotic Na+/H+ exchanger [Escherichia coli] sp P12703 XICE_ECOLI HYPOTHETICAL 60.5 KD PROTEIN IN SOXK-ACS NTERGENIC REGION (0549)	59	30	744
664	1	566	285	gi11262748	[LAMP-PV like component [Staphylococcus aureus]	59	33	282
670	1	3	455	gi1122758	[unknown [Bacillus subtilis]	59	42	453
674	3	543	929	gi1293033	[integrase [Bacteriophage phi-LC3]	59	46	387
758	1	349	176	gi11500472	[M. jannaschii predicted coding region M21577 [Methanococcus jannaschii]	59	37	174
771	2	2270	1461	gi1522150	[bromoperoxidase BPO-A1 [Streptomyces aureofaciens] sp P3912 BPAL_STRAU NON-HAEM BROMOPEROXIDASE BPO-A1 (EC 3.11.1.-) BROMIDE PEROXIDASE (BPO1).	59	44	810
825	1	2191	1097	gi1197526	[clumping factor [Staphylococcus aureus]	59	47	1095
1052	2	1094	723	gi1289262	[cane ORF3 [Bacillus subtilis]	59	36	372
1152	1	373	188	gi11276668	[ORE218 gene product [Porphyra purpurea]	59	37	186

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	START (nt)	STOP (nt)	Match position	Match gene name	% sim	% ident	length (nt)
1198	1	492	247	gi 142439	ATP-dependent nuclease [Bacillus subtilis]	59	26	246
1441	1	468	235	gi 1045942	glycyl-cRNA synthetase [Mycoplama genitalium]	59	37	234
2103	1	1	386	gi 459250	triacylglycerol lipase [Galactomyces geotrichum]	59	33	186
2205	1	793	398	gi 1303794	Yqem [Bacillus subtilis]	59	38	396
2578	1	484	284	gi 258003	insulin-like growth factor binding protein complex acid-labile subunit [rats, liver, Peptide, 603 aa]	59	48	201
2967	2	145	348	gi 1212730	Yqhk [Bacillus subtilis]	59	44	204
3012	1	3	248	gi 1773571	neurofilament protein NF70 [Helix asperae]	59	31	246
3544	1	3	401	gi 1055218	crotonase [Clostridium acetobutylicum]	59	42	399
3548	1	3	401	gi 1055218	crotonase [Clostridium acetobutylicum]	59	42	399
3560	1	698	351	gi 1055218	crotonase [Clostridium acetobutylicum]	59	42	348
3720	1	722	363	gi 1408494	homologous to penicillin acylase [Bacillus subtilis]	59	36	360
4171	1	3	296	gi 1055218	crotonase [Clostridium acetobutylicum]	59	42	294
4305	1	618	310	gi 1524193	unknown [Mycobacterium tuberculosis]	59	39	309
18	1	1242	622	gi 146913	N-acetylglucosamine transport protein [Escherichia coli] pir B29895 MOBC2N phosphotransferase system enzyme II [EC 7.1.69] N-acetylglucosamine-specific - Escherichia coli gp P09123 PTAA-ECOLI PTS SYSTEM, N-ACETYLGLUCOSAMINE-SPECIFIC IIABC ORPHANT (ETIA)	58	43	621
20	7	7020	5845	gi 50502	collagen alpha chain precursor (AA 27 to 1127) [Mus musculus]	58	50	1170
21	5	7234	3626	gi 1034860	phosphotibosyl aminotriazole isomerase [Thermotoga maritima]	58	32	393
21	2	2841	1669	gi 1276880	EpsC [Streptococcus thermophilus]	58	29	1173
23	10	9301	8090	pir A31133 A311	diaminopimelate decarboxylase [EC 4.1.1.20] - Pseudomonas aeruginosa	58	37	1212
38	29	22555	22884	gi 973249	vestibular reductase [Medicago sativa]	58	37	330
44	1	2	406	gi 289272	ferrichrome-binding protein [Bacillus subtilis]	58	33	405
45	1	1	552	gi 29464	embryonic myosin heavy chain (1085 AA) [Homo sapiens] tr S12460 S12460 myosin beta heavy chain - human	58	33	552
55	2	759	538	gi 1158852	glucose regulated protein [Echinococcus multilocularis]	58	32	222
62	13	8493	8068	gi 975353	kinase-associated protein B [Bacillus subtilis]	58	35	426
63	3	1553	1717	gi 164926	[Arabidopsis thaliana unidentified mRNA sequence, complete cds.]_ene product [Arabidopsis thaliana]	58	35	165

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
67	13	12017	11229	gi1228083	NADH dehydrogenase subunit 2 [Chorthippus parallelus]	58	41	789
96	8	8208	9167	gi1205992	hypothetical protein [Bacillus subtilis]	58	42	960
107	2	2065	1364	gi1806327	Escherichia coli hpaA gene for A protein similar to yeast PRP16 and RP22 [Escherichia coli]	58	37	702
112	7	4519	5613	gi1355568	glucose-fructose oxidoreductase [Zymomonas mobilis] pir[A42289]A42289 glucose-fructose oxidoreductase (EC 1.1.1.1) recursor - Zymomonas mobilis	58	38	1095
114	6	7118	6503	gi1337843	unknown [Bacillus subtilis]	58	38	816
143	2	2261	1395	gi1436051A436	mature-parasite-infected erythrocyte surface antigen HESA - Plasmodium falciparum	58	31	867
151	2	717	950	gi11370261	unknown [Mycobacterium tuberculosis]	58	31	234
154	6	6015	4627	gi11209277	pCTHm1 gene product [Chlamydia trachomatis]	58	41	1389
154	16	14281	13541	gi1146613	DNA ligase (EC 6.5.1.2) [Escherichia coli]	58	39	741
155	3	2269	1892	gi11303917	YqjB [Bacillus subtilis]	58	34	378
174	1	1056	539	gi1904198	hypothetical protein [Bacillus subtilis]	58	26	528
189	4	1533	1769	gi1467383	DNA binding protein (probable) [Bacillus subtilis]	58	25	237
201	3	2669	3307	gi11511453	endonuclease III [Methanococcus jannaschii]	58	34	639
208	1	2	238	gi11276729	phycobilisome linker polypeptide [Porphyra purpurea]	58	29	237
220	11	14575	13058	gi1397526	clumping factor [Staphylococcus aureus]	58	51	1518
231	3	1629	1474	gi11002520	HutS [Bacillus subtilis]	58	45	156
233	6	4201	3497	gi11463023	No definition line found [Caenorhabditis elegans]	58	39	705
243	10	9303	10082	gi1537207	ORF_277 [Escherichia coli]	58	32	780
257	1	331	1143	gi11340128	ORF1 [Staphylococcus aureus]	58	44	813
302	2	460	801	gi146174	ORF X [Bacillus subtilis]	58	34	342
307	11	6984	6127	gi11303842	YqjU [Bacillus subtilis]	58	30	858
321	3	1914	2747	gi11219996	hypothetical protein [Bacillus subtilis]	58	41	834
342	4	2724	3497	gi1454838	ORF 6: putative [Pseudomonas aeruginosa]	58	41	774
348	1	1	663	gi1467478	unknown [Bacillus subtilis]	58	36	663
401	2	384	605	gi1143407	para-aminobenzoic acid synthase, component 1 (pab) [Bacillus subtilis]	58	53	222

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
437	1	125	1556	gi11301866	YngS [Bacillus subtilis]	58	35	1210
445	1	105	1442	gi1581581	protein A [Staphylococcus aureus]	58	32	1338
453	3	789	965	gi11009455	unknown [Schizosaccharomyces pombe]	58	34	177
453	5	2748	2047	gi1537214	yjgC gene product [Escherichia coli]	58	40	702
479	2	731	1444	gi11256621	26.7% of identity in 165 aa to a Thermophilic bacterium hypothetical protein 6; putative [Bacillus subtilis]	58	36	714
490	1	909	547	gi1580920	rodD (gtAA) polypeptide (AA 1-673) [Bacillus subtilis] pI[506048]SD6048 probable rod protein - Bacillus subtilis sp P13484 TAGL_BACSU PROBABLE POLY(GLYCEROL-PHOSPHATE) LPHAS-GLUCOSYLTRANSFERASE (EC 2.4.1.52) (TECHNOIC ACID BIOSYNTHESIS PROTEIN E)	58	36	363
517	1	1	1164	gi1472641018	HYPOTHETICAL HELICASE M0018	58	30	1164
517	6	4182	4544	gi1453422	orf268 gene product [Mycoplasma hominis]	58	29	363
546	3	2402	4019	gi1884052	restriction modification system S subunit (Spiroplasma citri) gi1884052 restriction modification system S subunit (Spiroplasma citri)	58	37	1218
562	1	3	179	gi143831	infS protein (AA 1-400) [Klebsiella pneumoniae]	58	34	177
600	2	1147	1156	gi11181819	unknown [Pseudomonas aeruginosa]	58	44	192
604	2	1231	1001	gi11001353	hypothetical protein [Synecocystis sp.]	58	41	231
619	1	1	504	gi1903748	integral membrane protein [Homo sapiens]	58	43	504
625	1	2	364	gi11208474	hypothetical protein [Synecocystis sp.]	58	43	363
635	1	1492	755	gi11510995	transaldolase [Methanococcus jannaschii]	58	41	718
645	1	1	846	gi1677882	ileal sodium-dependent bile acid transporter [Rattus norvegicus] gi1677882 ileal sodium-dependent bile acid transporter [Rattus norvegicus]	58	33	846
645	3	906	1556	gi11239999	hypothetical protein [Bacillus subtilis]	58	41	651
665	1	771	532	gi11204262	hypothetical protein (CB:U1028-61) [Haemophilus influenzae]	58	39	240
674	1	635	327	gi1498817	ORF8: homologous to small subunit of phage terminases [Bacillus subtilis]	58	39	309
675	2	1312	806	gi142181	osmC gene product [Escherichia coli]	58	28	507
745	1	618	310	gi11205432	coenzyme PQQ synthase protein III (pqqIII) [Haemophilus influenzae]	58	32	309
799	2	242	1174	gi11204669	collagenase [Haemophilus influenzae]	58	36	933
800	2	1096	614	gi1171963	tRNA isopentenyl transferase [Saccharomyces cerevisiae] sp P07884 H005_YEAST TRNA ISOPENTENYLTRANSFERASE (EC 2.5.1.8) ISOPENTENYL-DIPHOSPHATE: TRNA ISOPENTENYLTRANSFERASE (IPP TRANSFERASE) (IPPT)	58	37	483

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	Sim	Ident	Length (nt)
854	1	1108	605	gi1466778	lysine specific peptidase (Escherichia coli)	58	44	504
865	1	481	242	gi1861199	protoporphyrin IX Mg-chelatase subunit precursor (Hordeum vulgare)	58	33	240
891	1	3	527	gi1291660	AbaA2 (Streptomyces coelicolor)	58	31	525
942	1	931	467	gi1405567	trmH (Pseudomonas DSM4181)	58	30	465
1002	1	952	521	gi1577649	preLUMK (Staphylococcus aureus)	58	34	432
1438	1	1	261	gi1581558	ISOLEUCYL-TRNA SYNTHETASE (Staphylococcus aureus) sp P41368 SVIP_STA001 ISOLEUCYL-TRNA SYNTHETASE, MUPIROICIN RESISTANT EC 6.1.1.5 (ISOLEUCINE--TRNA LIGASE) (ILERS) (MUPIROICIN RESISTANCE PROTEIN)	58	30	261
1442	1	2	463	gi1971194	similar to Acc No. D26185 (Escherichia coli)	58	34	462
1873	1	480	241	gi11319951	small subunit of NADH dependent glutamate synthase (Plectononema boryanum)	58	38	240
1876	1	3	158	gi1529216	No definition line found (Caenorhabditis elegans) sp P46503 YUX7_CAEZL HYPOTHETICAL 7.3 KD PROTEIN F23F12.7 IN HR230506 III	58	33	156
1989	1	108	401	gi11405458	YneR (Bacillus subtilis)	58	29	294
2109	1	3	401	gi11001801	hypothetical protein (Synechocystis sp.)	58	31	399
2473	1	288	145	gi1510140	lipoendopeptidase F (Lactococcus lactis)	58	38	144
2523	1	452	228	gi1644873	catabolic dehydroquinase dehydratase (Acinetobacter calcoaceticus)	58	37	225
3041	1	2	211	gi11205367	oligopeptide transfer ATP-binding protein (Haemophilus influenzae)	58	39	210
3094	1	3	263	gi11185288	isochorismate synthase (Bacillus subtilis)	58	38	261
3706	1	3	383	gi1416614	mevalonate kinase (Arabidopsis thaliana)	58	48	381
3854	1	1	402	gi1808869	human gcp372 (Homo sapiens)	58	32	402
4062	1	51	224	gi1508551	ribulose-1,5 biphosphate carboxylase large subunit-methyltransferase (Pisum sativum)	58	37	174
4278	1	3	206	gi1180189	cerebellar-degeneration-related antigen (CDR14) (Homo sapiens) gi1182737 cerebellar degeneration-associated protein (Homo sapiens) pir A29770 A29770 cerebellar degeneration-related protein - human	58	37	204
19	7	7818	7363	gi11001516	hypothetical protein (Synechocystis sp.)	57	31	456
23	11	9663	8872	gi1608066	ORF_F256 (Escherichia coli)	57	29	792
31	1	4801	2402	gi1153146	ORF3 (Streptomyces coelicolor)	57	32	2400
38	14	11611	10796	gi1144859	ORF B (Clostridium perfringens)	57	31	816
46	14	112063	11046	gi11001319	hypothetical protein (Synechocystis sp.)	57	25	984

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
51	3	1411	1187	pir B33856 B338	hypothetical 80K protein - <i>Bacillus sphaericus</i>	57	38	225
54	3	1	453	gi 684950	staphylococcal accessory regulator A ( <i>Staphylococcus aureus</i> )	57	31	453
75	3	3	239	gi 1000470	C2787.7 ( <i>Caenorhabditis elegans</i> )	57	42	237
92	5	3855	3061	gi 143607	sporulation protein ( <i>Bacillus subtilis</i> )	57	35	795
96	3	4006	4773	gi 144297	acetyl esterase (XyNC) ( <i>Caldocellum saccharolyticum</i> ) pir B37202 B37202	57	34	768
					acetyl esterase (EC 3.1.1.6) (XyNC) - <i>Caldocellum saccharolyticum</i>			
107	3	1460	2076	gi 460955	TagE ( <i>Vibrio cholerae</i> )	57	42	597
109	8	5340	5933	gi 1438846	Unknown ( <i>Bacillus subtilis</i> )	57	41	594
112	9	6679	7701	gi 1466250	Unknown ( <i>Bacillus subtilis</i> )	57	33	1023
114	4	6384	8108	gi 871456	putative alpha subunit of formate dehydrogenase ( <i>Mechanobacterium hercynotrophicum</i> )	57	37	2277
126	2	430	1053	gi 288301	ORF2 gene product ( <i>Bacillus megaterium</i> )	57	37	624
131	5	6537	6277	gi 1511160	M. jamae-like predicted coding region MJ1163 ( <i>Mechanococcus jamae</i> )	57	38	261
133	3	2668	2201	gi 1101912	Yqhw ( <i>Bacillus subtilis</i> )	57	40	468
133	4	3383	2784	gi 1221884	(urea?) amidolyase ( <i>Haemophilus influenzae</i> )	57	37	600
147	4	2164	1694	gi 467469	Unknown ( <i>Bacillus subtilis</i> )	57	33	471
160	2	1293	1060	gi 558604	chitin synthase 2 ( <i>Neurospora crassa</i> )	57	28	234
163	8	5687	4764	gi 145580	carD gene product ( <i>Escherichia coli</i> )	57	38	924
168	6	4336	5325	gi 39782	33kDa lipoprotein ( <i>Bacillus subtilis</i> )	57	32	990
170	5	3297	3455	gi 603404	Yer164p ( <i>Saccharomyces cerevisiae</i> )	57	37	159
221	6	8026	6809	gi 1136221	carboxypeptidase ( <i>Sulfolobus solfataricus</i> )	57	32	1218
228	3	1348	1791	gi 288969	fibronectin binding protein ( <i>Streptococcus dysgalactiae</i> ) pir S33850 S33850	57	32	444
					fibronectin-binding protein - <i>Streptococcus dysgalactiae</i>			
263	4	4411	3686	gi 1185002	dihydrodipicolinate reductase ( <i>Pseudomonas syringae</i> pv. <i>tabaci</i> )	57	42	726
276	1	494	255	gi 396380	No definition line found ( <i>Escherichia coli</i> )	57	40	240
283	2	335	1324	gi 773349	BlrA protein ( <i>Bacillus subtilis</i> )	57	32	990
297	1	469	236	gi 1334820	reading frame V ( <i>Cauliflower mosaic virus</i> )	57	46	234
342	3	1993	2805	gi 1204431	hypothetical protein (SP-P33644) ( <i>Haemophilus influenzae</i> )	57	35	813

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Protein ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
375	6	3340	3741	gi1385177	cell division protein [Bacillus subtilis]	57	26	402
433	6	3286	4011	gi13524117	alpha-acetolactate decarboxylase [Lactococcus lactis]	57	40	726
470	3	903	1145	gi1404819	protein serine/threonine kinase [Toxoplasma gondii]	57	30	243
487	5	1391	1723	gi1507323	ORF1 [Bacillus stearothermophilus]	57	28	333
498	1	274	852	gi1334549	NADH-ubiquinone oxidoreductase subunit 4L [Podospore anserina]	57	34	579
503	1	343	173	gi13502283	organic cation transporter OCT2 [Rattus norvegicus]	57	30	171
505	2	1619	1284	gi1466884	B1496_C2_194 [Mycobacterium leprae]	57	40	336
519	2	1182	2549	gi1303707	YvxA [Bacillus subtilis]	57	34	1368
522	2	3234	1945	gi13064809	homologous to sp.HTRA_ECO1 [Bacillus subtilis]	57	36	1290
538	2	909	1415	gi1153179	phosphorinothricin N-acetyltransferase [Streptomyces coelicolor] pirJ0246 [J0246 phosphinothricin N-acetyltransferase (EC 2.3.1.-)] Streptomyces coelicolor	57	40	507
547	1	968	486	gi1467340	unknown [Bacillus subtilis]	57	50	483
599	1	1062	532	gi1420692 [TYRA_	PHENYLAMINE DEHYDROGENASE (EC 1.3.1.12) (PDH)	57	41	531
620	2	757	572	gi1107894	unknown [Schizosaccharomyces pombe]	57	38	186
622	2	1600	1130	gi1173028	thioredoxin II [Saccharomyces cerevisiae]	57	39	471
625	2	362	1114	gi11262366	hypothetical protein [Mycobacterium leprae]	57	34	753
680	1	1	204	gi1143544	RNA polymerase sigma-30 factor [Bacillus subtilis] pirJ28625 [A28625 transcription initiation factor sigma H - actillus subtilis]	57	30	204
690	1	3	629	gi1466520	pocr [Salmonella typhimurium]	57	29	627
696	1	2	433	gi1413972	ipa-48r gene product [Bacillus subtilis]	57	33	432
704	1	36	638	gi1169931	M. jamaeensis predicted coding region M21083 [Methanococcus jamaeensis]	57	36	603
732	1	2316	1621	gi11418999	orf6 [Lactobacillus sakei]	57	37	696
746	1	451	227	gi1392973	Rub3 [Aplysia californica]	57	42	225
757	1	20	466	gi143979	L. curvatus small cryptic plasmid gene for rep protein [Lactobacillus curvatus]	57	45	447
862	1	2	295	gi1303827	YqfI [Bacillus subtilis]	57	21	294
1049	1	907	455	gi11510108	ORF-1 [Agrobacterium tumefaciens]	57	35	453
1117	1	1387	695	gi1896286	NR2 terminal uncertain [Leishmania tarentolae]	57	28	693

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	Accession	Ident	Length (nt)
1136	1	2	322	gi1303853	YnfF (Bacillus subtilis)		57	321
1144	2	1013	611	gi1310081	voltage-activated calcium channel alpha-1 subunit (Rattus norvegicus)		57	421
1172	1	1472	738	gi1511146	H. jennaschii predicted coding region MJ1143 (Methanococcus jennaschii)		57	735
1500	2	746	558	gi142780	putative membrane protein; putative (Bacillus subtilis)		57	189
1676	1	659	399	gi1313777	uracil permease (Escherichia coli)		57	261
2481	1	2	400	gi1237015	OMP4 (Bacillus subtilis)		57	399
3099	1	3	230	gi1204540	isochlorogenic synthase (Haemophilus influenzae)		57	228
3122	1	360	381	gi1882472	OMP-0464 (Escherichia coli)		57	180
3560	1	2	361	gi153490	tetracycline C resistance and export protein (Streptomyces laevis)		57	360
3850	1	856	414	gi1155588	glucose-fructose oxidoreductase (Zymomonas mobilis) pfrA42289 glucose-fructose oxidoreductase (EC 1.1.1.8) recursor - Zymomonas mobilis		57	423
3931	1	704	354	gi1413953	ape-29d gene product (Bacillus subtilis)		57	351
3993	1	1	384	gi151259	HMG-CoA reductase (EC 1.1.1.88) (Pseudomonas nevalonii) pfrA44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.		57	384
4065	1	793	398	gi1300377	nitrate reductase (EC 1.7.99.4) alpha chain - Escherichia coli		57	398
4100	1	586	300	gi1086633	T06C10.5 gene product (Caenorhabditis elegans)		57	297
4163	1	571	287	gi121512	patatin (Solanum tuberosum)		57	285
4267	2	631	335	gi1000165	SpotIIAG (Bacillus subtilis)		57	287
4358	1	3	302	gi1298032	IEF (Streptococcus suis)		57	300
4389	2	108	290	gi1405894	1-phosphofructokinase (Escherichia coli)		57	183
4399	1	2	232	gi11483603	pristinamycin I synthase I (Streptomyces pristinaespiralis)		57	231
4481	1	572	288	gi1405879	yeaH (Escherichia coli)		57	285
4486	1	512	258	gi1515938	glutamate synthase (ferredoxin) (Synochocystis sp.) pfrS46957 glutamate synthase (ferredoxin) (EC 1.4.7.1) - Synochocystis sp.		57	255
4510	1	481	242	gi1205301	leukotoxin secretion ATP-binding protein (Haemophilus influenzae)		57	240
4617	1	468	256	gi1511222	restriction modification enzyme, subunit M1 (Methanococcus jennaschii)		57	253
4	11	12201	11524	gi149204	histidine utilization repressor G (Klebsiella aerogenes) pfrA36730 hutG protein - Klebsiella pneumoniae (fragment) sp19452HUTG-KLEAE FORMINOGUTAMATE (EC 3.5.3.8) FORMINOGUTAMATE HYDROLASE (HISTIDINE UTILIZATION PROTEIN G) FRAGMENT		56	678

TABLE 2



S. aureus - putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
22	8	4248	5177	gi 1322222	RACH1 (Homo sapiens)	56	33	930
38	28	21179	22264	gi 1480705	lipote-protein ligase (Mycoplasma capricolum)	56	34	1086
44	3	1861	2421	gi 490120	Y gene product (unidentified)	56	31	561
44	15	10103	10606	gi 1205099	hypothetical protein (GB119201.1) (Haemophilus influenzae)	56	39	504
50	6	4820	5161	gi 209331	fiber protein (human adenovirus type 5)	56	48	342
51	4	2076	2972	gi 623476	transcriptional activator (Providencia stuartii) sp P43463 AARP_PROST TRANSCRIPTIONAL ACTIVATOR AARP	56	30	897
67	6	5856	6594	gi 466613	nikB (Escherichia coli)	56	32	939
89	3	2364	1810	gi 482922	protein with homology to pal: repressor of B. subtilis (Lactobacillus elbrusckii)	56	39	555
96	1	203	913	gi 145594	cAMP receptor protein (cpx) (Escherichia coli)	56	35	711
109	21	18250	17846	gi 1204367	hypothetical protein (GB14003.278) (Haemophilus influenzae)	56	27	405
112	8	5611	6678	gi 155588	glucose-fructose oxidoreductase (Zymomonas mobilis) pfr A42289 A42289 glucose-fructose oxidoreductase (LC 1.1.-.-) recursor - Zymomonas mobilis	56	40	1068
131	3	6404	5100	gi 619724	MgC (Bacillus firmus)	56	30	1305
138	2	65	232	gi 413948	ipa-24d gene product (Bacillus subtilis)	56	31	168
138	4	823	1521	gi 580868	ipa-22r gene product (Bacillus subtilis)	56	31	699
146	2	740	447	gi 1046009	M. genitalium predicted coding region MG309 (Mycoplasma genitalium)	56	37	294
149	2	1619	1067	gi 945180	trnminase small subunit (Bacteriophage LU-III)	56	35	573
163	1	2	223	gi 143947	glutamine synthetase (Bacteroides fragilis)	56	30	222
166	5	6745	6449	gi 405792	ORF154 (Pseudomonas putida)	56	26	297
187	1	31	393	gi 311237	H(+)-transporting ATP synthase (Zea mays)	56	30	363
190	1	2	373	gi 1109686	ProX (Bacillus subtilis)	56	35	372
191	8	11518	9943	gi 561070	acyl coenzyme A synthetase (Escherichia coli)	56	35	1596
195	1	1291	647	gi 1510242	collagenase (Methanococcus jannaschii)	56	34	645
230	3	2323	2072	gi 40363	heat shock protein (Clostridium acetobutylicum)	56	39	252
238	5	3383	3775	gi 1477513	serA (Staphylococcus aureus)	56	31	393
270	2	813	1712	gi 765073	autolysin (Staphylococcus aureus)	56	41	900

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
290	1	3221	1632	gi 547513	orf3 [Haemophilus influenzae]	56	34	1570
297	5	1140	1373	gi 1511556	M. jannaschii predicted coding region MJ1561 [Methanococcus jannaschii]	56	40	234
321	2	2947	1799	gi 1001801	hypothetical protein [Synecocystis sp.]	56	31	1149
359	2	1279	641	gi 46336	ncf1 gene product [Rhizobium meliloti]	56	26	639
371	2	360	1823	gi 145304	L-ribulokinase [Escherichia coli]	56	39	1464
391	4	1762	2409	gi 1061634	hypothetical protein [Synecocystis sp.]	56	34	648
402	1	380	192	gi 1418904	5-HT4L receptor [Homo sapiens]	56	48	189
416	4	2480	2109	gi 1408486	HS74A gene product [Bacillus subtilis]	56	31	372
424	3	1756	2334	gi 142471	acetolactate decarboxylase [Bacillus subtilis]	56	32	579
457	1	1907	1017	gi 1205194	formaldehyde dehydrogenase [Haemophilus influenzae]	56	36	891
458	2	2423	1812	gi 15466	terminase [Bacteriophage SP11]	56	37	612
504	2	2152	1283	gi 1142681	lpp38 [Pasteurella haemolytica]	56	38	870
511	1	1	1284	gi 217049	brnQ protein [Salmonella typhimurium]	56	37	1284
601	3	1094	1701	gi 467109	rim, 30S ribosomal protein S16 alanine acetyltransferase, 229, CL170 [Mycobacterium leprae]	56	43	603
660	5	3547	3774	gi 1229706	28K910.1 [Caenorhabditis elegans]	56	30	228
707	1	35	400	gi 151929	NADPH-sulfite reductase flavoprotein component [Salmonella typhimurium]	56	38	366
709	2	1145	1095	gi 1510801	hydrogenase accessory protein [Methanococcus jannaschii]	56	38	291
718	1	1	495	gi 413948	ipa 24d gene product [Bacillus subtilis]	56	35	495
743	1	87	677	gi 928836	repressor protein [Lactococcus lactis phage BK5-7]	56	35	591
790	1	776	399	gi 1511513	ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii]	56	33	378
795	1	3	407	gi 1205382	cell division protein [Haemophilus influenzae]	56	34	405
813	1	19	930	gi 1222161	permease [Haemophilus influenzae]	56	28	912
855	1	3	515	gi 1256621	26.7% of identity in 165 aa to a Thermophilic bacterium hypothetical protein 6; putative [Bacillus subtilis]	56	33	513
968	1	2	466	gi 547513	orf3 [Haemophilus influenzae]	56	37	465
973	2	1049	732	gi 886022	HexR [Pseudomonas aeruginosa]	56	31	318
1203	1	5	223	gi 184251	HMG-1 [Homo sapiens]	56	34	219

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
1976	1	422	237	gi 19806	lysine-rich aspartic acid-rich protein [Plasmidium chabaudi] r[S22183]S22183 lysine/aspartic acid-rich protein - plasmidium baui	56	33	216
2161	1	2	400	gi 1237015	ORF4 [Bacillus subtilis]	56	27	399
2958	1	362	183	gi 466685	No definition line found [Escherichia coli]	56	26	180
2979	1	421	212	gi 1204354	spore germination and vegetative growth protein [Haemophilus influenzae]	56	40	210
2994	2	556	376	gi 816616	phosphoribosylformylino prae ketolomerase [Rhodobacter phaeoideus]	56	29	201
3026	1	179	328	gi 143306	penicillin V amidase [Bacillus sphaericus]	56	30	150
3189	1	289	146	gi 1166604	Similar to aldehyde dehydrogenase [Caenorhabditis elegans]	56	37	144
3770	1	63	401	gi 1129145	acetyl-CoA C-acyltransferase [Mangifera indica]	56	43	339
4054	2	720	361	gi 1205355	Na <sup>+</sup> /H <sup>+</sup> antiporter [Haemophilus influenzae]	56	31	360
4145	1	3	324	gi 1226095	long-chain acyl-CoA dehydrogenase [Mus musculus]	56	36	324
4200	1	505	254	gi 155588	glucose-fructose oxidoreductase [Zymomonas mobilis] pIR[A42289]A42289 glucose-fructose oxidoreductase [EC 1.1.1.-] recuadr - Zymomonas mobilis	56	40	252
4273	1	675	355	gi 308861	CTG start codon [Lactococcus lactis]	56	33	321
1	3	4095	3436	gi 5341	Putative orf YCC8C, len:192 [Saccharomyces cerevisiae] r[S53591]S53591 hypothetical protein - yeast [Saccharomyces cerevisiae]	55	25	660
11	12	9377	4505	gi 216773	haloacetate dehalogenase H-1 [Moraxella sp.]	55	32	473
12	4	5133	4534	gi 467337	unknown [Bacillus subtilis]	55	26	600
19	5	5404	5644	gi 1001719	hypothetical protein [Synecocystis sp.]	55	25	441
21	13	14087	12339	gi 474190	luca gene product [Escherichia coli]	55	30	1749
32	7	5164	6448	gi 1340096	unknown [Mycobacterium tuberculosis]	55	37	1321
34	3	2569	1808	gi 1303968	YqjQ [Bacillus subtilis]	55	39	762
34	5	3960	3412	gi 1303962	YqjK [Bacillus subtilis]	55	33	549
36	1	1291	647	gi 406045	ORF_018 [Escherichia coli]	55	27	645
36	6	6220	5243	gi 1001341	hypothetical protein [Synecocystis sp.]	55	31	978
47	3	3054	3821	gi 1001819	hypothetical protein [Synecocystis sp.]	55	21	768
49	1	2065	1127	gi 403373	glycerophosphoryl diester phosphodiesterase [Bacillus subtilis] pIR[S37251]S37251 glycerophosphoryl diester phosphodiesterase - actillus subtilis	55	36	939

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
67	11	8966	9565	gi 153053	norA199 protein [Staphylococcus aureus]	55	73	600
75	3	881	1273	gi 11698	L-histidinol, NAD: oxidoreductase (EC 3.1.1.23) [aa 1-436] [Escherichia coli]	55	31	393
82	9	15387	14194	gi 1136221	carboxypeptidase [Sulfolobus solfataricus]	55	35	1194
87	4	3517	4917	gi 1064812	[function unknown] [Bacillus subtilis]	55	26	1401
94	2	1172	1636	gi 182463	protein-Nipil-phosphohistidine-sugar phosphotransferase [Escherichia coli]	55	35	465
92	1	127	516	gi 1377832	[unknown] [Bacillus subtilis]	55	36	390
100	2	836	2035	gi 1370274	seaxanthin epoxidase [Nicotiana glauca]	55	36	1200
100	5	5137	4658	gi 396640	[unknown open reading frame] [Buchnera aphidicola]	55	29	480
104	3	4266	2986	gi 149866	M. jamaensis predicted coding region M1024 [Methanococcus jamaensis]	55	31	1281
116	3	2616	1834	gi 1511367	formate dehydrogenase, alpha subunit [Methanococcus jamaensis]	55	29	783
144	3	1805	1476	gi 1100787	[unknown] [Saccharomyces cerevisiae]	55	35	330
165	5	6212	5508	gi 1045884	M. genitalium predicted coding region M0199 [Mycoplasma genitalium]	55	27	705
189	5	2205	2576	gi 143569	ATP synthase alpha subunit [Bacillus firmus]	55	35	372
191	6	9136	6857	gi 559411	B0272.3 [Caenorhabditis elegans]	55	39	2280
194	2	364	636	gi 1145768	K7 kinesin-like protein [Dictyostelium discoideum]	55	34	273
209	4	1335	1676	gi 1473357	chi4 gene product [Schizosaccharomyces pombe]	55	35	342
211	2	1693	1145	gi 1410130	ORF6 [Bacillus subtilis]	55	37	549
213	2	644	1372	gi 633692	TrsA [Yersinia enterocolitica]	55	28	729
214	7	4144	5481	gi 1001793	hypothetical protein [Synecocystis sp.]	55	30	1338
221	7	11473	9197	gi 466520	pocR [Salmonella typhimurium]	55	32	2277
230	8	5908	4817	gi 1237063	[unknown] [Mycobacterium tuberculosis]	55	38	1092
236	4	1375	2340	gi 1146199	putative [Bacillus subtilis]	55	32	966
243	2	380	1885	gi 459907	mercuric reductase (plasmid p1250)	55	29	1506
258	1	786	394	gi 455006	orf6 [Rhodococcus fascians]	55	36	393
281	1	126	938	gi 1408493	homologous to SwissProt:YMDA_ECOLI: hypothetical protein [Bacillus subtilis]	55	35	811
316	3	1323	2102	gi 1486447	LuxA homologue [Rhizobium sp.]	55	30	780
326	5	2968	2744	gi 1296824	proline iminopeptidase [Lactobacillus helveticus]	55	36	225

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
351	2	2322	1429	gi11204820	hydrogen peroxide-inducible activator (Haemophilus influenzae)	55	24	894
353	4	2197	2412	gi11272475	chitin synthase (Emerella nidulans)	55	50	216
380	1	14	379	gi1142554	ATP synthase I subunit (Bacillus megaterium)	55	37	366
383	1	482	232	gi1149272	ferrichrome-binding protein (Bacillus subtilis)	55	36	231
386	1	3	938	gi11510251	DNA helicase, putative (Methanococcus jannaschii)	55	30	936
410	2	1208	1891	gi11205144	multidrug resistance protein (Haemophilus influenzae)	55	27	684
483	2	411	833	gi1113934	lipo-10r gene product (Bacillus subtilis)	55	26	423
529	3	1777	1433	gi1406150	ORF_1309 (Escherichia coli)	55	33	345
555	1	1088	585	gi1143407	para-aminobenzoic acid synthase, component I (pub) (Bacillus subtilis)	55	28	504
565	1	402	202	gi11223961	CDP-tylase epimerase (Yersinia pseudotuberculosis)	55	41	201
582	1	751	452	gi11256643	20.2% identity with NADH dehydrogenase of the Leishmania major mitochondrion; putative (Bacillus subtilis)	55	36	300
645	5	2260	2057	gi11410424	fusion protein F (Bovine respiratory syncytial virus) p1rJQ1481.V028A (fusion glycoprotein precursor - bovine respiratory syncytial virus (strain A51908))	55	25	204
672	2	957	2216	gi11511333	M. jannaschii predicted coding region M1322 (Methanococcus jannaschii)	55	36	1280
730	1	955	479	gi11537007	ORF_1379 (Escherichia coli)	55	30	472
737	1	1859	945	gi11536963	CG Site No. 18166 (Escherichia coli)	55	30	915
742	2	228	572	gi1104160	product unknown (Bacillus subtilis)	55	38	345
817	2	1211	903	gi11136289	histidine kinase A (Dictyostelium discoideum)	55	29	309
819	1	582	355	gi11558073	polymorphic antigen (Plasmodium falciparum)	55	22	228
832	2	1152	724	gi1140367	ORF1 (Clostridium acetobutylicum)	55	32	429
840	1	769	386	gi11205875	pseudouridylyl synthase I (Haemophilus influenzae)	55	39	384
1021	1	23	529	gi1148563	beta-lactonase (Yersinia enterocolitica)	55	38	507
1026	1	60	335	gi1147804	Opp C (MAL-301) (Salmonella typhimurium)	55	26	276
1525	1	1	282	gi11477533	serA (Staphylococcus aureus)	55	49	482
1814	2	224	985	gi11046078	H. genitalium predicted coding region HG369 (Mycoplasma genitalium)	55	38	762
3254	1	427	254	gi11413968	lipo-44d gene product (Bacillus subtilis)	55	30	174

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
3695	1	686	345	gi 216773	haloacetate dehalogenase H-1 [Moraxella sp.]	55	32	342
3721	1	1	312	gi 42029	ORF1 gene product [Escherichia coli]	55	31	312
3799	1	3	272	gi 42029	ORF1 gene product [Escherichia coli]	55	36	270
3889	1	22	423	gi 1129145	acetyl-CoA C-acyltransferase [Mangifera indica]	55	45	402
3916	1	2	385	gi 529754	apeC [Streptococcus pyogenes]	55	38	384
3945	1	4	198	gi 476252	phase 1 flagellin [Salmonella enterica]	55	36	195
4074	1	488	246	gi 42029	ORF1 gene product [Escherichia coli]	55	38	243
4184	1	2	343	gi 1524267	unknown [Mycobacterium tuberculosis]	55	28	342
4284	1	14	208	gi 1100774	ferredoxin-dependent glutamate synthase [Synecococcus sp.]	55	36	195
4457	2	644	378	gi 180189	cerebellar-degeneration-related antigen (CDR14) [Homo sapiens] gi 182737	55	38	267
4514	1	2	244	gi 216773	haloacetate dehalogenase H-1 [Moraxella sp.]	55	32	243
4599	1	432	217	gi 1129145	acetyl-CoA C-acyltransferase [Mangifera indica]	55	42	216
4606	1	416	210	gi 386120	myosin alpha heavy chain (S2 subfragment) [rabbits, masseter, epitope partial, 234 aa]	55	27	207
5	8	5348	4932	gi 536069	ORF YBL047c [Saccharomyces cerevisiae]	54	27	417
12	7	7166	6165	gi 1205504	homoserine acetyltransferase [Haemophilus influenzae]	54	30	1002
23	16	17086	15326	gi 474192	lucC gene product [Escherichia coli]	54	31	1761
35	1	2	979	gi 48054	small subunit of soluble hydrogenase (AA 1-384) [Synecococcus sp.]	54	36	978
37	11	9437	8667	gi 337207	ORF F277 [Escherichia coli]	54	38	771
37	12	8165	8132	gi 1160967	palmitoyl-protein thioesterase [Homo sapiens]	54	37	168
46	15	13025	13804	gi 438473	protein is hydrophobic, with homology to E. coli P10M, putative Bacillus subtilis	54	28	760
56	2	203	736	gi 1256139	YbbJ [Bacillus subtilis]	54	34	534
57	13	11117	10179	gi 1151248	inosine-uridine preferring nucleoside hydrolase [Crithidia fasciculata]	54	32	919
66	2	516	1133	gi 1335781	[Cap [Drosophila melanogaster]	54	29	618
70	10	8116	8646	gi 1399823	PhoE [Rhizobium meliloti]	54	31	531

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
70	15	12556	11801	gi1064811	TETRA-CYCLOPEPTIDE RESISTANCE PROTEIN	54	29	756
87	5	4915	5706	gi1064811	[function unknown (Bacillus subtilis)]	54	33	792
92	4	3005	2289	gi1205366	[oligopeptide transport ATP-binding protein (Haemophilus influenzae)]	54	33	717
103	2	2596	1556	gi1710495	[protein kinase (Bacillus brevis)]	54	33	1041
105	2	3585	2095	gi1143727	[putative (Bacillus subtilis)]	54	30	1491
112	4	2137	2712	gi153724	[HspC (Streptococcus pneumoniae)]	54	41	396
127	2	1720	2493	gi1144297	[acetyl esterase (XynC) (Caldococcus saccharolyticum) pir g37202 g37202 acetyl esterase (EC 3.1.1.6) (XynC) - Caldococcus saccharolyticum]	54	34	774
138	5	1600	3306	gi142473	[pyruvate oxidase (Escherichia coli)]	54	36	1707
152	2	525	1172	gi11377834	[unknown (Bacillus subtilis)]	54	23	648
161	9	4831	5469	gi1903305	[ORF73 (Bacillus subtilis)]	54	28	639
161	13	6694	7251	gi1511039	[phosphate transport system regulatory protein (Methanococcus jannaschii)]	54	32	558
164	6	1263	4543	gi1204976	[prolyl-tRNA synthetase (Haemophilus influenzae)]	54	34	1281
164	120	21602	32243	gi1143582	[apoptin protein (Bacillus subtilis)]	54	32	642
171	6	5683	4250	gi1438965	[malA gene products (Bacillus stearothermophilus) pir S43914 S43914 hypochlorite protein 1 - Bacillus stearothermophilus]	54	37	1434
206	18	19208	19720	gi1240016	[R09E10.3 (Caenorhabditis elegans)]	54	38	513
218	2	1090	1905	gi1467378	[unknown (Bacillus subtilis)]	54	26	816
220	1	1322	663	gi1353761	[myosin II heavy chain (Naegleria fowleri)]	54	22	660
220	113	12655	13059	pir S00485 S004	[gene 11-1 protein precursor - Plasmodium falciparum (fragments)]	54	35	405
221	3	2030	3709	gi11303813	[yqew (Bacillus subtilis)]	54	34	1680
272	7	5055	4219	gi162964	[arylamine N-acetyltransferase (AA 1-290) (Gallus gallus) ir S06652 NYCHY3 arylamine N-acetyltransferase (EC 2.3.3.5) (clone NAT-3) - chicken]	54	33	837
316	7	4141	4701	gi1682769	[eccE gene product (Escherichia coli)]	54	31	561
316	10	6998	8742	gi1413951	[ipa-27d gene product (Bacillus subtilis)]	54	28	1749
338	3	3377	2214	gi1450328	[LORF F (unidentified)]	54	28	1164
341	4	3201	3634	gi1171959	[myosin-like protein (Saccharomyces cerevisiae)]	54	25	414

TABLE 2

S aureus Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
346	1	1820	912	gi 396400	similar to eukaryotic Na <sup>+</sup> /H <sup>+</sup> exchangers [Escherichia coli] sp P32703 VJCE_ECOLI_HYPOTHEICAL 60.5 KD PROTEIN IN SOKR-ACS NTERGENIC REGION (0549)	54	34	909
348	2	623	1351	gi 537109	ORF f343a [Escherichia coli]	54	34	729
378	2	1007	1942	sp P02983 TCP_S	TETRACYCLINE RESISTANCE PROTEIN	54	31	936
408	6	4351	5301	gi 474190	lucA gene product [Escherichia coli]	54	29	951
444	9	7934	8854	gi 216267	ORF2 [Bacillus megaterium]	54	32	921
463	2	2717	2229	gi 304160	product unknown [Bacillus subtilis]	54	50	489
502	2	1696	1133	gi 1205015	hypothetical protein (SP:P10120) [Haemophilus influenzae]	54	38	564
505	6	6262	5357	gi 1500558	12-hydroxyhepta-2,4-diene-1,7-dioate isomerase [Methanococcus jannaschii]	54	41	906
550	1	2736	1522	gi 40100	ProC (tag3) polypeptide (AA 1-746) [Bacillus subtilis] ir S06049 S06049 rddc protein - Bacillus subtilis p P13485 TAGP_BACSU TECHOIC ACID BIOSYNTHESIS PROTEIN F	54	35	1215
551	5	3305	4279	gi 950197	unknown [Corynebacterium glucanum]	54	34	975
558	2	1356	958	gi 483090	No definition line found [Caenothobitis elegans]	54	32	399
580	1	91	936	gi 331906	[fused envelope glycoprotein precursor (Friend spleen focus-forming virus)]	54	45	846
603	3	554	757	gi 1323423	ORF YCR234w [Saccharomyces cerevisiae]	54	36	204
617	1	25	249	gi 219959	ornithine transcarbamylase [Homo sapiens]	54	40	235
622	3	1097	1480	gi 1301873	Yqo2 [Bacillus subtilis]	54	25	384
623	1	3	404	gi 1063250	low homology to P20 protein of Bacillus licheniformis and bleomycin acetyltransferase of Streptomyces verticillius [Bacillus subtilis]	54	45	402
689	1	1547	1011	gi 552446	NADH dehydrogenase subunit 4 [Apis mellifera liguatical pir S52968 S52968 NADH dehydrogenase chain 4 - honeybee itochondrion (SGC4)]	54	30	537
725	2	686	1443	gi 987096	sensory protein kinase [Streptomyces hygroscopicus]	54	26	756
956	1	1	249	pir S10782 S107	integrin homolog - yeast [Saccharomyces cerevisiae]	54	24	249
978	2	1137	859	gi 1301994	ORF YNL091w [Saccharomyces cerevisiae]	54	33	279
1314	1	3	281	gi 1001108	hypothetical protein [Synchocystis sp.]	54	33	279
2450	1	1	228	gi 1045057	ch-TGG [Homo sapiens]	54	32	228
2934	1	1	387	gi 580870	ipa-37d gxaA gene product [Bacillus subtilis]	54	36	387
2970	1	499	251	sp P1734P VECE	HYPOTHETICAL PROTEIN IN ASP5 5'-REGION (FRAGMENT)	54	42	249

TABLE 2



S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3002	1	1	309	gi 144027	Tme protein [Lactococcus lactis]	54	33	309
3561	1	9	464	gi 151259	HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas mvaloni] pir A44756 A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	54	15	466
3572	1	72	401	gi 450698	hcdM gene of Ecopri gene product [Escherichia coli] pir S38437 S38437 hcdM protein - Escherichia coli: pir S09629 S09629 hypothetical protein A - Escherichia coli (S08 40-520)	54	36	330
3829	1	798	400	gi 132245	mevalonate pyrophosphate decarboxylase [Rattus norvegicus]	54	29	399
3909	1	1	273	gi 29865	CENP-E (Homo sapiens)	54	30	273
3921	1	3	209	pir S24325 S243	glucan 1,4-beta-glucosidase (EC 3.2.1.74) - Pseudomonas fluorescens subsp. cellulosa	54	34	207
4438	1	566	285	gi 1196657	unknown protein [Mycoplasma pneumoniae]	54	30	282
4459	1	3	272	gi 1046081	hypothetical protein (CB:026185_10) [Mycoplasma genitalium]	54	38	270
4564	1	3	221	gi 216267	ORF2 [Bacillus megaterium]	54	38	219
23	12	12538	10685	gi 474192	lucC gene product [Escherichia coli]	53	35	1854
23	14	14841	13579	gi 42029	ORF1 gene product [Escherichia coli]	53	32	1263
24	3	4440	3940	gi 1369947	G2 gene product [Bacteriophage B1]	53	36	501
26	4	3818	4618	gi 1486247	unknown [Bacillus subtilis]	53	37	801
34	6	2856	3998	gi 405880	Yell [Escherichia coli]	53	40	1143
38	10	9380	7806	gi 139954	thyroid sodium/iodide symporter NIS [Rattus norvegicus]	53	29	1575
56	10	12324	12100	pir A54592 A545	110k actin filam associated protein - chicken	53	32	225
57	6	5047	4583	pir A00141 DE2P	alcohol dehydrogenase (EC 1.1.1.1) - fission yeast [Schizosaccharomyces pombe]	53	39	465
57	12	10515	8932	gi 1486429	putative transcriptional regulator [Bacillus stearothermophilus]	53	30	1584
67	12	9496	10218	gi 1511555	quinolone resistance norA protein protein [Methanococcus jamaehilli]	53	31	723
69	3	3125	2382	gi 1087017	acetabulogalactan-protein, AGP [Nicotiana glauca, cell-suspension culture filtrate, peptide, 461 aa]	53	30	744
79	1	3	1031	gi 1523802	glucanase [Anabaena variabilis]	53	32	1029
80	1	673	338	gi 452428	ATPase 3 [Plasmodium falciparum]	53	36	336
88	4	1910	2524	gi 137034	ORF_0488 [Escherichia coli]	53	25	615
88	5	2467	3282	gi 137034	ORF_0488 [Escherichia coli]	53	29	816

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match	Accession	Match gene name	% sim	% ident	length (nt)
92	6	5870	5505	gi 359598		amphotropic murine retrovirus receptor (Rattus norvegicus)	53	33	366
94	5	4417	3239	gi 173018		tropomyosin (TPM) (Saccharomyces cerevisiae)	53	25	1179
99	5	4207	5433	sp P28246 BCR_L		BICYCLOMYCIN RESISTANCE PROTEIN (SULFONAMIDE RESISTANCE PROTEIN)	53	30	1227
120	3	1639	2262	gi 576655		ORF1 (Vibrio anguillarum)	53	35	624
120	11	7257	8897	gi 1524397		glycine betaine transporter Opd (Bacillus subtilis)	53	33	1641
127	6	6883	5685	gi 1256630		putative (Bacillus subtilis)	53	32	1209
147	2	255	557	gi 581648		lepin gene product (Staphylococcus epidermidis)	53	34	303
154	4	4705	4256	gi 151004		mucoidy regulatory protein Algr (Pseudomonas aeruginosa) p1r A32802 A32802 regulatory protein algr - Pseudomonas aeruginosa sp P26275 ALGR_PSEAE POSITIVE ALGINATE BIOSYNTHESIS REGULATORY PROTEIN	53	32	450
171	7	5717	5421	gi 1510669		hypothetical protein (GP:D64044_18) (Methanococcus jannaschii)	53	34	287
191	9	13087	11483	gi 298085		acetate decarboxylase (Clostridium acetobutylicum) p1r B49346 B49346 butyrate-acetate CoA-transferase (EC 8.3.9) small chain - Clostridium acetobutylicum sp P3752 CTFA_CLOB BUTYRATE-ACETATE COA-TRANSFERASE SUBUNIT (EC 2.8.3.9) (COAT A)	53	31	1605
201	5	3763	4326	gi 143456		IrpoE protein (ttg start codon) (Bacillus subtilis)	53	29	564
206	17	18204	18971	gi 304136		acetylglutamate kinase (Bacillus stearothermophilus) sp 007905 ARGB_BACST ACETYLGLUTAMATE KINASE (EC 2.7.2.8) (NAG INASE) (AGK) (N-ACETYL-L-GLUTAMATE 5-PHOSPHOTRANSFERASE)	53	36	768
212	10	4021	4271	gi 19878		protein kinase (Plasmodium falciparum)	53	28	201
231	2	1580	1350	gi 537506		paramyosin (Dirofilaria immitis)	53	34	231
272	6	2719	3249	gi A31141 A311		hypothetical protein (gtfD 3' region) - Streptococcus mutans	53	34	531
308	3	927	2576	gi 1606292		ORF_0696 (Escherichia coli)	53	33	1650
320	7	5645	5884	gi 160596		RNA polymerase III largest subunit (Plasmodium falciparum) sp P27625 RPCL_PFAFA DNA-DIRECTED RNA POLYMERASE III LARGEST SUBUNIT (EC 2.7.7.6)	53	33	240
327	1	218	901	gi 854601		unknown (Schizosaccharomyces pombe)	53	31	684
341	2	212	2500	gi 633712		ORF1 (Campylobacter jejuni)	53	31	2289
351	1	763	383	gi P1675 YABM		HYPOTHETICAL 42.7 KD PROTEIN IN THPA-LDUD INTERGENIC REGION (ORF104)	53	32	381
433	7	5087	4731	gi 1001961		MHC class II analog (Staphylococcus aureus)	53	30	357
454	2	1240	980	gi A60128 A603		40K cell wall protein precursor (sr 5' region) - Streptococcus mutans (strain OH2175, serotype f)	53	27	261

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Astq gene name	% sim	% ident	length (nt)
470	4	1123	1761	gi 516826	rat GCP360 (Rattus rattus)	53	30	639
483	1	432	217	gi 1480439	putative transcriptional regulator (Bacillus stearothermophilus)	53	23	216
544	1	516	1259	gi 46587	ORF 1 (AA 1 - 121) (1 is 2nd base in codon) (Staphylococcus aureus) tr S15765 S15765 hypothetical protein 1 (nib 5' region) - aphyllococcus aureus (fragment)	53	38	744
558	10	3957	3754	gi 15140	res gene (Bacteriophage P1)	53	32	204
603	2	319	620	gi 507738	Hmp (Vibrio parahaemolyticus)	53	26	282
693	1	1669	941	gi 153123	toxic shock syndrome toxin-1 precursor (Staphylococcus aureus) pir A24606 KCSAS1 toxic shock syndrome toxin-1 precursor - taphylococcus aureus	53	38	729
766	1	2	673	gi 687600	orfA2: orfA2 forms an operon with orfA1 (Listeria monocytogenes)	53	43	672
781	1	667	335	gi 1204551	pilin biogenesis protein (Haemophilus influenzae)	53	26	333
801	1	1	545	gi 1279400	ISAP protein (Escherichia coli)	53	25	543
803	1	2	910	gi 695278	lipase-like enzyme (Alcaligenes eutrophus)	53	30	909
872	1	1177	590	gi 298032	EF (Streptococcus suis)	53	30	588
910	1	2	184	gi 1044936	unknown (Schistosoma mansoni)	53	29	183
943	1	794	399	gi 190508	similar to unidentified ORF near 47 minutes (Escherichia coli) sp P31436 VICK_ECOLI HYPOTHETICAL 43.5 KD PROTEIN IN SELC-NLPA NTGENIC REGION	53	30	396
988	1	1004	504	gi 142441	ORF 3: putative (Bacillus subtilis)	53	28	501
1064	1	3	434	gi 305080	myosin heavy chain (Entamoeba histolytica)	53	26	432
1366	1	3	452	gi 308852	transmembrane protein (Lactococcus lactis)	53	33	450
1758	1	792	397	gi 1001774	hypothetical protein (Synecocystis sp.)	53	30	396
1897	1	1	447	gi 1303949	Yqix (Bacillus subtilis)	53	27	447
2381	1	798	400	gi 1146243	22.4% identity with Escherichia coli DNA-damage inducible protein putative (Bacillus subtilis)	53	37	399
3537	1	1	327	gi 450688	hsdM gene of EcoRI gene product (Escherichia coli) pir S18437 S18437 hsdM protein - Escherichia coli pir S09629 S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	53	35	327
3747	2	137	397	gi 1477486	transposase (Burkholderia cepacia)	53	53	261
11	5	3049	3441	gi 868224	No definition line found (Caenorhabditis elegans)	52	33	393

TABLE 2

S. aureus - putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
15	5	2205	2369	gi 215966	G41 protein (gag start codon) [Bacteriophage T4]	52	34	165
19	3	2429	3808	gi 1205379	UDP-murac-pentapeptide synthetase (Haemophilus influenzae)	52	31	1380
24	1	6920	3462	gi 579124	predicted 86.4kd protein; 52kd observed [Mycobacteriophage L5] p1r[S10971]S10971 gene 26 protein - Mycobacterium phage L5 ap[Q05233]UG26_RPM15 MINOR TAIL PROTEIN GP26. (SUB 2-837)	52	32	3659
37	5	3015	3935	gi 1500543	P115 protein [Methanococcus jannaschii]	52	25	921
38	13	8795	9703	gi 46851	glucose kinase [Streptomyces coelicolor]	52	29	909
44	16	10617	11066	gi 42012	ImoA gene product [Escherichia coli]	52	36	450
46	1	3	521	gi 1040957	NADH dehydrogenase subunit 6 [Anopheles trinkae]	52	25	519
51	10	5531	6280	gi 388269	trac [Plasmodium falciparum]	52	32	750
56	5	3964	2826	gi 181949	endothelial differentiation protein (edg-1) [Homo sapiens] p1r[A35300]A35300 G protein-coupled receptor edg-1 - human ap[P2145]EDG1_HUMAN PROBABLE G PROTEIN-COUPLED RECEPTOR EDG-1.	52	23	1143
57	5	4850	4173	gi 304153	sorbitol dehydrogenase [Bacillus subtilis]	52	27	678
62	5	3364	2870	gi 1072399	phae gene product [Rhizobium meliloti]	52	25	495
62	6	4445	3651	gi 146485	NADH dehydrogenase [Synecococcus PCC7942]	52	27	795
67	14	11355	12962	gi 1511365	glutamate synthase (NADPH), subunit alpha [Methanococcus jannaschii]	52	30	1608
67	21	16935	18158	gi 1204393	hypothetical protein (SP P1122) [Haemophilus influenzae]	52	25	1224
70	4	2185	1997	gi 7227	cytoplasmic dynein heavy chain [Dictyostelium discoideum] r[A4357]A4357 dynein heavy chain, cytosolic - a1me mold cytosolium discoideum	52	36	189
96	10	10005	10664	gi 1408485	B65G gene product [Bacillus subtilis]	52	26	660
103	5	3986	3351	gi 11009368	respiratory nitrate reductase [Bacillus subtilis]	52	42	636
109	3	4102	3350	gi 599274	labe gene product [Mycobacterium leprae]	52	39	753
109	19	15732	17300	gi 1526981	amino acid permease YeeF like protein [Salmonella typhimurium]	52	30	1569
121	3	1412	981	gi 772931	unknown [Saccharomyces cerevisiae]	52	32	432
125	3	865	1680	gi 1286975	put gene product [Porphyromonas gingivalis]	52	38	816
130	2	659	1807	gi 1256634	25-8% identity over 120 aa with the Synecococcus sp. PCC7942 [Bacillus subtilis]	52	36	1149
149	1	1164	583	gi 1225943	P85X terminase [Bacillus subtilis]	52	33	582
149	14	4687	4415	gi 1510368	M. jannaschii predicted coding region M0272 [Methanococcus jannaschii]	52	35	273

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match	match gene name	% sim	% ident	length (nt)
167	1	276	1001	gi1146025		cell division protein [Escherichia coli]	52	43	786
188	1	120	1256	gi1474915		orf 337, translated orf similarity to SW: BCR-ECOLI bicyclomycin resistance protein of Escherichia coli [Coxiella burnetii] pir[S44207][S44207] hypothetical protein 337 - Coxiella burnetii (SUS - 338)	52	26	1137
195	9	9161	8760	gi11028		mitochondrial outer membrane 72K protein [Neurospora crassa] r[A3682][A3682 72K mitochondrial outer membrane protein - rompura crassa]	52	25	402
200	3	2065	2607	gi1142439		ATP-dependent nuclease [Bacillus subtilis]	52	35	543
203	4	2776	3684	gi11303698		hcd [Bacillus subtilis]	52	25	909
227	8	5250	5651	gi1305080		myosin heavy chain [Entamoeba histolytica]	52	24	402
242	1	21	1424	gi11060877		Emv [Escherichia coli]	52	32	1404
249	5	4526	4753	pir[C3722][C372		cytochrome P450 1A1, hepatic - dog (fragment)	52	23	228
255	1	2107	1055	gi1143290		penicillin-binding protein [Bacillus subtilis]	52	28	1053
276	7	3963	3664	gi11001610		hypothetical protein [Synecococcus sp.]	52	10	300
276	8	4456	4055	gi1146235		orf L3 [Mycoplasma capricolum]	52	26	402
289	2	1856	1449	gi1150900		GTP phosphohydrolase [Proteus vulgaris]	52	34	408
325	1	1	279	gi11204874		polypeptide deformylase (formylmethionine deformylase) [Haemophilus influenzae]	52	33	279
340	1	2017	1010	gi11215695		peptide transport system protein SapF homolog, SapF homolog [Mycoplasma pneumoniae]	52	33	1008
375	3	340	1878	gi1467446		similar to SpvB [Bacillus subtilis]	52	28	1539
424	4	4104	3262	gi11478239		unknown [Mycobacterium tuberculosis]	52	34	843
430	1	3	575	pir[A42606][A426		orfA 5, to orf405 - Saccharopolyspora erythraea (fragment)	52	28	573
444	4	4728	3712	gi11408494		homologous to penicillin acylase [Bacillus subtilis]	52	31	1017
465	1	1802	903	gi1143331		alkaline phosphatase regulatory protein [Bacillus subtilis] pir[A27650][A27650 regulatory protein phoR - Bacillus subtilis sp P23545 PHO_R BACSU ALKALINE PHOSPHATASE SYNTHESIS SENSOR PROTEIN PHO (BC 2.7.3.-)]	52	36	900
469	5	4705	4169	gi1755152		highly hydrophobic integral membrane protein [Bacillus subtilis] sp P42953 TAGG-BACSU TEICHOIC ACID TRANSLOCATION PERMEASE PROTEIN AGG	52	32	537
495	1	1262	633	gi11204607		transcription activator [Haemophilus influenzae]	52	25	630
505	7	6004	5762	gi1142440		ATP-dependent nuclease [Bacillus subtilis]	52	28	243

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
517	2	1162	1614	gi 166162	Bacteriophage phi-11 int gene activator [Staphylococcus aureus] phage phi 11	52	35	453
543	2	444	1295	gi 1215693	putativeorf: G9_0rf434 [Mycoplasma pneumoniae]	52	25	852
586	1	1	336	gi 581648	epib gene product [Staphylococcus epidermidis]	52	36	336
771	1	848	426	gi 1279769	EdHC [Methanobacterium thermoformicicum]	52	30	423
1120	2	100	330	gi 142439	ATP-dependent nuclease [Bacillus subtilis]	52	35	231
1614	1	691	347	gi 289262	comE ORF [Bacillus subtilis]	52	28	345
2495	1	1	324	gi 216151	DNA polymerase (gene L, tlg start codon) [Bacteriophage SP02] gi 579197 SP02 DNA polymerase (aa 1-648) [Bacteriophage SP02] pif A21458 [UBPS2 DNA-directed DNA polymerase (EC 2.7.7.7) - phage P02]	52	34	324
2931	1	566	285	gi 1256136	YbBG [Bacillus subtilis]	52	30	282
2943	1	577	320	gi 41713	hnsA ORF (AA 1-245) [Escherichia coli]	52	35	258
2993	1	588	295	gi 298032	EP [Streptococcus suis]	52	34	294
3667	1	612	307	gi 849025	hypothetical 64.7-kDa protein [Bacillus subtilis]	52	36	306
3944	1	478	260	gi 1218040	BAA [Bacillus licheniformis]	52	36	219
3954	2	613	347	gi 854064	ORF [Human herpesvirus 6]	52	50	267
3946	1	90	401	gi 1205919	Na <sup>+</sup> and Cl <sup>-</sup> dependent gamma-aminobutyric acid transporter [Haemophilus influenzae]	52	33	312
4002	1	3	389	gi 40003	oxoglutarate dehydrogenase (NADP+) [Bacillus subtilis] p p23125 [OD01_BACSU 2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT (EC 2.4.2) ALPHA-KETOGLUTARATE DEHYDROGENASE]	52	42	387
4020	1	1	249	gi 159388	ornithine decarboxylase [Leishmania donovani]	52	47	249
4098	1	438	220	gi 409795	No definition line found [Escherichia coli]	52	32	219
4248	1	3	212	gi 965077	Adrep [Saccharomyces cerevisiae]	52	40	210
7	1	3	575	gi 895747	putative cel operon regulator [Bacillus subtilis]	51	28	571
21	4	2479	3276	gi 1510962	indole-3-glycerol phosphate synthase [Methanococcus jannaschii]	51	32	798
22	9	5301	5966	gi 1303933	YqjN [Bacillus subtilis]	51	25	666
43	3	1516	1283	gi 1519460	Srp1 [Schizosaccharomyces pombe]	51	31	234
44	17	11042	11305	gi 42011	moed gene product [Escherichia coli]	51	35	264
51	11	6453	6731	gi 495471	vacuolating toxin [Helicobacter pylori]	51	37	279

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
52	4	2537	2995	gi1256652	25% identity to the E.coli regulatory protein MprA, putative [Bacillus subtilis]	51	32	459
57	10	7231	6843	gi1508173	EIIA domain of PTS-dependent Gal transport and phosphorylation Escherichia coli	51	32	489
59	1	9	1111	gi1299163	alanine dehydrogenase [Bacillus subtilis]	51	33	1083
67	20	15791	16576	gi1510977	M. jannaeschii predicted coding region M0938 [Methanococcus jannaeschii]	51	24	786
69	2	1559	1218	gi1467359	unknown [Bacillus subtilis]	51	34	342
71	1	3	1196	gi1298032	rf [Streptococcus suis]	51	32	1194
78	2	349	176	gi11161242	proliferating cell nuclear antigen [Styela clava]	51	28	174
99	4	3357	4040	gi1642795	TFIID subunit TAF155 (Homo sapiens)	51	25	684
109	1	2852	1428	gi1580920	rodd (gtaA) polypeptide (AA 1-673) [Bacillus subtilis] pir[S06048]S06048 probable rodd protein - Bacillus subtilis sp[P13484]TAGE BACSU PROBABLE POLYGLYCEROL-PHOSPHATE) LPHA-GLUCOSYLTRANSFERASE (EC 2.4.1.52) (TECHNOIC ACID BIOSYNTHESIS PROTEIN E)	51	27	1425
109	9	6007	6693	gi11204815	hypothetical protein (SP:P32662) [Haemophilus influenzae]	51	23	687
112	3	1066	2352	pir[S05330]S053	maltose-binding protein precursor - Enterobacter aerogenes	51	42	1287
112	13	18432	12855	gi1405857	yehU [Escherichia coli]	51	29	1578
114	9	9725	8967	gi1435098	orf1 [Mycoplasma capricolum]	51	30	759
115	1	1	912	gi11431110	[OMP YML095w [Saccharomyces cerevisiae]	51	28	912
127	10	9647	10477	gi11204314	H influenzae predicted coding region H10056 [Haemophilus influenzae]	51	37	831
152	9	6814	7356	gi1431929	Mun1 regulatory protein [Mycoplasma sp.]	51	38	543
154	2	575	1153	gi11237044	unknown [Mycobacterium tuberculosis]	51	36	579
154	7	6587	5634	gi1409286	bmrU [Bacillus subtilis]	51	27	954
171	8	6943	6236	gi11205484	hypothetical protein (SP:P33918) [Haemophilus influenzae]	51	32	708
184	1	1	291	gi1466886	bl496_C3_206 [Mycobacterium leprae]	51	33	291
212	5	1501	2139	pir[A45605]A456	mature-parasite-infected erythrocyte surface antigen HESA - Plasmodium falciparum	51	23	639
228	2	707	1378	gi14204	nuclear protein [Drosophila melanogaster]	51	27	672
236	8	8137	7481	gi149272	Aspergillase [Bacillus licheniformis]	51	31	657
243	4	4637	3546	gi11511102	melvalonate kinase [Methanococcus jannaeschii]	51	29	1092

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match position	Match gene name	% sim	% ident	length (nt)
257	4	3540	3373	gi 204579	H. Influenzae predicted coding region H10326 [Haemophilus influenzae]	51	22	168
258	3	2397	1609	gi 160299	glutamic acid-rich protein [Plasmodium falciparum] pfr A54514 A54514 glutamic acid-rich protein precursor - Plasmodium aliciparum	51	34	789
265	5	2419	3591	gi 580841	PI [Bacillus subtilis]	51	32	1173
298	2	518	748	gi 1336162	SCP8 [Streptococcus agalactiae]	51	34	231
316	9	5817	7049	gi 413953	ipa-29d gene product [Bacillus subtilis]	51	39	1233
332	2	3775	2057	gi 1209012	mutS [Thermus aquaticus thermophilus]	51	26	1719
364	4	3816	4991	gi 528991	unknown [Bacillus subtilis]	51	32	1176
440	2	448	684	gi 2819	transferase (GAL10) (AA 1 - 687) [Kluyveromyces fragilis] r S01407 KXUWG UDPGlucose 4-epimerase (EC 5.1.3.2) - yeast uyeromyces marxianus var. lactis	51	32	231
495	2	1353	1177	gi 297861	protease G [Ewingia chrysanthemi]	51	41	177
495	3	2267	1718	gi 1513317	isatine rich protein [Entamoeba histolytica]	51	25	570
506	1	840	421	gi 455320	icII protein [Bacteriophage P4]	51	33	420
600	1	1474	983	gi 587532	orf. len: 201. CAT: 0.16 [Saccharomyces cerevisiae] pfr S48818 S48818 hypothetical protein - yeast [Saccharomyces erevisiae]	51	30	492
607	3	479	934	gi 1511524	hypothetical protein (SP-P37002) [Methanococcus jannaschii]	51	60	456
686	2	127	600	gi 493017	endocarditis specific antigen [Enterococcus faecalis]	51	30	474
726	1	33	230	gi 1353851	unknown [Prochlorococcus marinus]	51	45	198
861	1	176	652	gi 410145	dehydroquinase dehydratase [Bacillus subtilis]	51	34	477
869	1	782	393	gi 40100	rodC (tag) polypeptide (AA 1-746) [Bacillus subtilis] ir S06049 S06049 rodC protein - Bacillus subtilis p P13485 TAGC_BACSU TECHN AC: D BIOSYNTHESIS PROTEIN F.	51	23	390
1003	1	642	322	gi 1279707	hypothetical phosphoglycerate mutase [Saccharomyces cerevisiae]	51	39	321
1046	2	866	624	gi 510257	glycosyltransferase [Escherichia coli]	51	29	243
1467	1	702	352	gi 1511175	M. jannaschii: predicted coding region M3177 [Methanococcus jannaschii]	51	32	351
2558	1	457	230	ap P10582 DPM	DNA POLYMERASE (EC 2.7.7.7) [S-1 DNA ORF 3]	51	26	228
3003	1	779	399	gi 1809543	CbrC protein [Ewingia chrysanthemi]	51	27	381
3604	1	1	399	pt JC4210 JC42	[3-hydroxyacyl-CoA dehydrogenase (EC 1.1.1.35) - mouse	51	37	399
3732	1	2	316	gi 145906	acyl-CoA synthetase [Escherichia coli]	51	33	315

TABLE 2



S. aureus Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
3991	1	2	274	gi11061351	semaphorin III family homolog (Homo sapiens)	51	37	273
3995	1	46	336	gi1236346	surfactin synthetase (Bacillus subtilis)	51	38	291
4193	1	612	307	gi142749	ribosomal protein L12 (AA 1-179) [Escherichia coli] ir[S04776]XKECP peptide N-acetyltransferase (EC 2.3.1.-) - chetichia coli	51	25	306
4539	1	367	185	gi11408494	homologous to penicillin acylase (Bacillus subtilis)	51	60	183
4562	1	432	239	gi11458280	coded for by C. elegans cDNA cm047; Similar to hydroxymethylglutaryl-CoA synthase (Caenorhabditis elegans)	51	35	204
1	4	3576	4859	gi1559160	GRAIL score: null; Cap site and late promoter motifs present; putative [Autographa californica nuclear polyhedrosis virus]	50	44	1284
11	7	4044	5165	gi11146207	putative (Bacillus subtilis)	50	35	1122
11	13	30309	9496	gi11208451	hypothetical protein (Synecocystis sp.)	50	39	1014
19	1	2034	1018	gi1413966	ipa-42d gene product (Bacillus subtilis)	50	29	1017
20	11	8586	8407	gi11323159	ORF YOR103w [Saccharomyces cerevisiae]	50	28	180
24	5	5408	4824	gi1496280	structural protein [Bacteriophage Tuc2009]	50	29	585
34	4	1926	2759	gi11303966	YqjO (Bacillus subtilis)	50	36	834
38	30	22865	23440	gi11072179	Similar to dihydroflavonol-4-reductase (maize, petunia, tomato) [Caenorhabditis elegans]	50	32	576
47	2	1705	2976	gi1153015	FemA protein (Staphylococcus aureus)	50	29	1272
56	13	15790	15841	gi1806096	ORF f167; end overlaps end of o100 by 14 bases; start overlaps t174, ther starts possible [Escherichia coli]	50	10	562
57	1	2135	1077	gi1440922	xylytol dehydrogenase (unidentified hemiascomycete)	50	29	1059
58	2	628	1761	gi1143725	putative (Bacillus subtilis)	50	29	1134
98	6	4393	3884	gi11072179	Similar to dihydroflavonol-4-reductase (maize, petunia, tomato) [Caenorhabditis elegans]	50	32	510
89	5	3700	3356	gi11276658	ORF174 gene product (Pisiphya purpurea)	50	25	345
141	1	3	239	gi1476024	carbamoyl phosphate synthetase II (Plasmodium falciparum)	50	33	237
151	1	186	626	gi11403441	unknown [Mycobacterium tuberculosis]	50	35	441
166	7	11065	9623	gi1895747	putative cel operon regulator (Bacillus subtilis)	50	32	1443
201	6	5284	5096	gi1160229	circumsporozoite protein (Plasmodium relictum)	50	42	189
206	22	30784	29555	gi11052754	Larp integral membrane protein (Lactococcus lactis)	50	24	1230

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
211	4	1523	1927	gi 410131	ORFX7 [Bacillus subtilis]	50	29	405
214	4	2411	3295	sp P37348 YECE_	HYPOTHETICAL PROTEIN IN ASP5 5' REGION (FRAGMENT)	50	37	885
228	7	5068	4406	gi 313580	envelope protein (human immunodeficiency virus type 1) p1r[S35835][S35835 envelope protein - human immunodeficiency virus type 1 (fragment) (SUS 1-77)]	50	35	663
272	2	3048	1723	gi 1408485	B65G gene product [Bacillus subtilis]	50	22	1326
273	2	1616	984	gi 194386	phosphoglycerate mutase [Saccharomyces cerevisiae]	50	28	613
338	2	2507	1605	gi 148896	lipoprotein [Haemophilus influenzae]	50	26	903
332	4	5469	3802	gi 1526547	DNA polymerase family X [Thermus aquaticus]	50	27	1668
342	5	3473	3931	gi 456562	G-box binding factor [Dictyostellum discoideum]	50	35	459
352	1	1478	741	gi 288301	ORF2 gene product [Bacillus megaterium]	50	29	738
408	7	5259	5523	gi 11465	ORF2136 [Marchantia polymorpha]	50	27	225
420	3	650	1825	gi 1757842	UDP-sugar hydrolase [Escherichia coli]	50	30	1176
464	1	1	591	gi 487282	Na <sup>+</sup> -ATPase subunit J [Enterococcus hirae]	50	29	591
472	2	1418	864	gi 551875	BglR [Lactococcus lactis]	50	23	555
520	1	23	541	gi 567036	CapE [Staphylococcus aureus]	50	27	519
529	1	6	410	gi 1256652	25% identity to the E. coli regulatory protein MprA, putative [Bacillus subtilis]	50	34	405
536	5	7726	6059	gi 295671	selected as a weak suppressor of a mutant of the subunit AC40 of DNA dependent RNA polymerase I and III [Saccharomyces cerevisiae]	50	18	1608
647	1	2990	1497	gi 405568	Tral protein shares sequence similarity with a family of oplosomereses [Plasmodium falciparum]	50	31	1494
664	3	1133	711	gi 410007	Leukocidin F component [Staphylococcus aureus MRSA No. 4, Peptide, 23 aa]	50	32	423
678	1	1	627	gi 298032	EF [Streptococcus suis]	50	29	627
755	3	947	1171	gi 150572	cytochrome c1 precursor (EC 1.10.2.2) [Paracoccus denitrificans] gi 45465 cytochrome c1 (AA 1-450) [Paracoccus denitrificans] p1r[C2941][C29413 ubiquinol:--cytochrome-c reductase (EC 1.10.2.2) cytochrome c1 precursor [Paracoccus denitrificans sp 13627 CV1]	50	37	225
827	1	1363	683	gi 142020	heterocyst differentiation protein (Anabaena sp.)	50	21	681
882	1	3	752	gi 478485	B65G gene product [Bacillus subtilis]	50	27	750
910	2	418	887	gi 1294727	tyrosine-specific transport protein [Haemophilus influenzae]	50	25	450

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
933	1	524	gi1205451	cell division inhibitor [Haemophilus influenzae]	50	32	237
973	1	424	gi1486947	orf3 gene product [Saccharomyces cerevisiae]	50	40	189
1009	1	653	gi1153727	M protein [group B streptococcus]	50	28	255
1027	1	511	gi141934	ipa 10r gene product [Bacillus subtilis]	50	25	255
1153	2	556	gi1771676	ncxa [Alcaligenes xylosoxydans]	50	36	231
1222	1	798	gi1408485	165G gene product [Bacillus subtilis]	50	21	399
1350	1	682	gi1288272	[ferrichrome-binding protein [Bacillus subtilis]	50	32	294
2945	1	366	gi1171704	hexaprenyl pyrophosphate synthetase (COG) [Saccharomyces erevisiae]	50	34	183
2968	2	1604	gi1397526	clumping factor [Staphylococcus aureus]	50	33	801
2998	2	657	gi1495696	PS4E7.3 gene product [Caenorhabditis elegans]	50	40	264
3046	2	506	gi1531819	acyl carrier protein - Anabaena variabilis (fragment)	50	32	201
3063	1	547	gi1174190	lucA gene product [Escherichia coli]	50	29	273
3174	1	5	gi1151900	alcohol dehydrogenase [Rhodospirillum rubrum]	50	31	144
3192	1	625	gi11001423	hypothetical protein [Synchocystis sp.]	50	35	312
3400	1	2	gi1144733	RNA-dependent beta hydroxybutyryl coenzyme A dehydrogenase [Clostridium acetobutylicum]	50	28	261
3946	1	373	gi1576765	cytochrome b [Myrmecia pilosula]	50	38	146
3984	1	578	gi137348	HYPOTHETICAL PROTEIN IN ASP5 5-REGION (FRAGMENT)	50	37	288
37	10	8250	gi1204367	hypothetical protein [GB:U14001.278] [Haemophilus influenzae]	49	30	366
46	16	13802	gi1466860	acd: B1308.F1.34 [Mycobacterium leprae]	49	24	1047
59	5	2267	gi1606304	ORF_0462 [Escherichia coli]	49	27	1335
112	18	17884	gi1559502	ND4 protein (AA 1 - 409) [Caenorhabditis elegans]	49	25	732
138	9	6971	gi1303953	esterase [Acinetobacter calcoaceticus]	49	29	930
217	6	4401	gi1496254	fibronectin/fibrinogen-binding protein [Streptococcus pyogenes]	49	31	738
220	12	11803	gi1397526	clumping factor [Staphylococcus aureus]	49	31	855
228	4	1842	gi1523592	hypothetical protein 9 - Plasmodium falciparum	49	24	651
268	1	5016	gi143047	ORF8 [Bacillus subtilis]	49	26	2403

TABLE 2

TABLE 2

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
271	2	1164	1373	gi 1001257	hypothetical protein [Synecocystis sp.]	49	38	210
300	3	4340	3180	gi 1510796	hypothetical protein (GP_X91006.2) [Methanococcus jannaschii]	49	26	1161
381	1	2281	1142	gi 396301	[matches E50004: Bacterial regulatory proteins, arcC family signature [Escherichia coli]]	49	29	1140
466	1	3	947	gi 103863	YqgP [Bacillus subtilis]	49	26	945
666	1	379	191	gi 633112	GRF1 [Streptococcus sobrinus]	49	29	189
670	2	403	1014	gi 1122758	unknown [Bacillus subtilis]	49	32	612
709	1	1433	795	gi 143830	YpAC [Bacillus subtilis]	49	29	639
831	1	943	473	gi 401786	phosphomannomutase [Mycoplasma pirum]	49	29	471
1052	1	422	213	gi 1301799	Yqen [Bacillus subtilis]	49	21	210
1800	1	342	172	gi 216300	[peptidoglycan synthesis enzyme [Bacillus subtilis] sp P37583 HURG_BACSU MURG PROTEIN UPD-N-ACETYLGLUCOSAMINE--N-ACETYLUMURAMYL- PENTAPEPTIDE PYROPHOSPHORYL-UNDECAPENOL N-ACETYLGLUCOSAMINE RANSFERASE]	49	28	171
2430	1	2	376	sp P27034 YECA	HYPOTHETICAL 36.2 KD PROTEIN IN NDK-UCPE INTERGENIC REGION.	49	26	375
3096	1	542	273	gi 516360	[surfactin synthetase [Bacillus subtilis]]	49	25	270
32	4	3771	3100	gi 1217963	[hepatocyte nuclear factor 4 gamma [HNF4gamma] [Homo sapiens]]	48	36	672
38	1	1	609	gi 205790	[H influenzae predicted coding region HI1555 [Haemophilus influenzae]]	48	28	609
45	6	5021	6427	gi 1574267	unknown [Mycobacterium tuberculosis]	48	20	1407
59	14	116346	31096	gi 1197336	[lmpJ protein [Mycoplasma hominis]]	48	28	14751
61	1	3	608	gi 1511555	[quinolone resistance nraA protein protein [Methanococcus jannaschii]]	48	30	606
61	3	3311	3646	gi 1503893	[YqNL [Bacillus subtilis]]	48	29	336
114	1	98	415	gi 671708	[suts] homolog, similar to Drosophila melanogaster suppressor of able [suts]] protein, Swiss-Prot Accession Number P22293 [Irsinophila virilis]	48	25	318
121	1	1131	610	gi 1314584	unknown [Sphingomonas S8W]	48	29	522
136	1	2014	1280	gi 1205968	[H. influenzae predicted coding region HI1738 [Haemophilus influenzae]]	48	21	735
171	10	8220	9557	gi 1208454	[hypothetical protein [Synecocystis sp.]	48	34	1338
175	1	3625	1814	gi 396400	[similar to eukaryotic Na+/H+ exchangers [Escherichia coli] sp P32703 YJCE_ECOLI HYPOTHETICAL 60.5 KD PROTEIN IN SOMA-ACS INTERGENIC REGION (0549)]	48	29	1812
194	1	2	385	gi 11510493	[M jannaschii: predicted coding region M30419 [Methanococcus jannaschii]]	48	25	384

S. aureus - putative coding regions of novel proteins similar to known proteins

Gene ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
137	1	901	452	gi11045714	spermidine/putrescine transport ATP-binding protein [Mycoplasma genitalium]	48	25	450
203	1	1	396	gi1940288	protein localized in the nucleoli of pea nuclei. ORF, putative Pisum sativum	48	29	396
204	1	1363	698	gi1529202	No definition line found [Caenorhabditis elegans]	48	25	666
206	20	14815	27760	gi1511490	gramicidin S synthetase 2 [Bacillus brevis]	48	27	7056
212	1	2	166	gi1295899	nucleolin [Xenopus laevis]	48	34	165
223	110	12652	11426	gi1144073	SecY protein [Lactococcus lactis]	48	23	1227
241	6	6450	5491	gi11184118	mevalonate kinase [Methanobacterium thermoautotrophicum]	48	30	960
264	4	5434	3308	gi11015903	ORF YJR151c [Saccharomyces cerevisiae]	48	26	2127
441	1	1512	768	gi1142863	replication initiation protein [Bacillus subtilis] pir[B26580]B26580	48	23	765
					replication initiation protein - Bacillus subtilis			
444	5	3498	5298	gi1145836	putative [Escherichia coli]	48	24	1401
484	2	388	1110	gi1146551	transmembrane protein (Kdpp) [Escherichia coli]	48	16	723
542	3	1425	2000	gi1528969	N-carbamoylserine amidohydrolase (EC 3.5.1.59) - Arthrobacter sp.	48	27	576
566	1	1	1019	gi1151490	tetracycline C resistance and export protein [Streptomyces laevis]	48	24	1017
611	1	2	730	gi11103507	unknown [Schizosaccharomyces pombe]	48	38	729
624	1	1255	665	gi1144859	ORF B [Clostridium perfringens]	48	26	591
846	1	1014	508	gi1537506	paraoxylin [Dirofilaria immitis]	48	27	507
1020	1	66	950	gi11499476	magnesium and cobalt transport protein [Methanococcus jannaschii]	48	30	845
1227	1	1	174	gi11493730	lipoxigenase [Pisum sativum]	48	35	174
1266	1	1	405	gi11862452	ORF_F211; alternate name yggA, orf5 of X14836 [Escherichia coli] gi11425	48	24	405
					ORF5 (AA 1-197) [Escherichia coli] (SUB 15-211)			
2071	1	707	381	gi11408486	HS74A gene product [Bacillus subtilis]	48	25	327
2198	1	463	233	gi11500401	reverse gyrase [Methanococcus jannaschii]	48	40	231
2425	1	476	246	gi1148563	GI protein - fowlpox virus (strain HP444) (fragment)	48	40	231
2432	1	446	225	gi11353703	Trilo [Homo sapiens]	48	33	227
2453	1	794	399	gi1142850	division initiation protein [Bacillus subtilis]	48	29	396
2998	1	469	236	gi11577565	PepV [Lactobacillus delbrueckii]	48	31	234

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
3042	1	14	280	gi 945219	leucin (Homo sapiens)	48	35	267
3686	1	1	405	gi 145836	putative [Escherichia coli]	48	25	405
4027	2	492	301	gi 551177 SS11	trans-activator protein - Equine infectious anemia virus	48	32	192
4	2	3641	2232	gi 1303989	YqkI [Bacillus subtilis]	47	24	1410
24	2	599	1084	gi 1540083	PC4-1 gene product [Bradyella hydral]	47	28	486
36	10	7524	6925	gi 1420923	esterase [Acinetobacter wolffii]	47	26	600
43	2	196	1884	gi 15403455	unknown [Mycobacterium tuberculosis]	47	27	1689
44	22	16118	15108	gi 1511555	quinolone resistance norA protein protein [Methanococcus jannaschii]	47	31	1011
69	7	7141	6710	gi 1438466	Possible operon with orfG. Hydrophilic, no homologue in the database. putative [Bacillus subtilis]	47	29	412
81	4	5022	4279	gi 1466882	bps1: B1096_C2_189 [Mycobacterium leprae]	47	24	744
120	12	9135	8863	gi 1927340	D9509.27p: CA1: 0.12 [Saccharomyces cerevisiae]	47	38	273
142	1	2022	1174	gi 1486143	ORF YKL094w [Saccharomyces cerevisiae]	47	32	849
168	1	2178	1093	gi 1117254	hypothetical Ec8B protein [Bacillus subtilis]	47	29	1086
264	1	1884	943	gi 1142822	O-alanine racemase cdb [Bacillus subtilis]	47	34	942
279	1	1109	561	gi 1516608	2 predicted membrane helices, homology with B. subtilis men Orf3 Rouland et. al. unpublished Accession number H71833, approximately 1 minutes on updated Rudi map; putative [Escherichia coli] sp P7355 VFBB_ECOLI HYPOTHETICAL 26.7 KD PROTEIN IN MEND-MENB	47	31	549
345	2	2620	1676	gi 1204835	hippocase [Haemophilus influenzae]	47	28	945
349	2	152	400	gi 1456562	G-box binding factor [Dictyostellum discoideum]	47	32	249
391	1	1	831	gi 11420856	myo-inositol transporter [Schizosaccharomyces pombe]	47	19	831
404	3	2072	2773	gi 1255425	C3308.2 gene product [Caenorhabditis elegans]	47	17	702
529	5	2145	3107	gi 1303973	XqIV [Bacillus subtilis]	47	29	963
565	2	2321	1257	gi 142824	processing protease [Bacillus subtilis]	47	28	1065
654	1	962	483	gi 1243353	ORF 5' of ECRF3 [herpesvirus salmisi IVS, host-aquittet monkey, eptide, 407 aa]	47	23	480
692	1	115	633	gi 150756	40 kDa protein [Plasmid pJMI]	47	25	519
765	1	1634	819	gi 1256621	26.7% of identity in 165 aa to a Thermophilic bacterium hypothetical protein 6, putative [Bacillus subtilis]	47	24	816

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
825	2	211	1023	gi1197526	clumping factor (Staphylococcus aureus)	47	32	813
914	1	1	615	gi1558073	polymorphic antigen (Plasmodium falciparum)	47	29	615
1076	1	1	753	gi1147557	Aspartate aminotransferase (Bacillus circulans)	47	33	753
1351	1	793	398	gi1755153	ATP-binding protein (Bacillus subtilis)	47	20	396
4192	1	3	253	gi145836	putative (Escherichia coli)	47	24	291
5	6	4704	4361	gi105080	myosin heavy chain (Leishmania histolytica)	46	30	348
11	4	2777	3058	gi1603639	Yel040p (Saccharomyces cerevisiae)	46	28	282
46	11	10518	10300	gi1246901	ATP-dependent DNA ligase (Candida albicans)	46	28	213
61	4	3941	7930	gi1298032	EF (Streptococcus suis)	46	35	3973
132	4	5028	4093	gi11511057	hypothetical protein SP:45869 (Methanococcus jannaschii)	46	25	936
170	4	4719	3652	pir1551910 S519	G4 protein - Sauroleishmania tarentolae	46	26	1064
191	7	9543	8284	gi1041334	F5405.7 (Caenorhabditis elegans)	46	25	1263
253	1	1	396	gi1204449	dihydrolysoamide acetyltransferase (Haemophilus influenzae)	46	35	396
264	3	437	973	gi180189	cerebellar degeneration-related antigen (CDR34) (Homo sapiens) gi1182737	46	29	517
				pir129770 A29770	cerebellar degeneration-related protein - human			
273	1	485	285	gi1607573	envelope glycoprotein (C2V) region (Human immunodeficiency virus type 1)	46	35	201
350	1	3	563	gi1537052	ORF_1286 (Escherichia coli)	46	35	561
404	1	2	862	gi11221884	(urea?) amidolyase (Haemophilus influenzae)	46	31	861
410	4	1876	2490	gi1110518	proton antiporter efflux pump (Mycobacterium smegmatis)	46	24	615
432	1	2663	1455	gi1197634	orf4: putative transporter; Method: conceptual translation supplied by author (Mycobacterium smegmatis)	46	27	1209
454	1	2419	1211	gi15470	portal protein (Bacteriophage SPPI)	46	30	1209
517	5	2477	4192	gi1523812	orf5 (Bacteriophage A2)	46	23	1716
540	3	1512	1285	gi1215635	pacA (Bacteriophage P1)	46	30	228
587	2	649	1242	gi1537148	ORF_181 (Escherichia coli)	46	29	594
1218	1	747	391	gi1205456	single-stranded-DNA-specific exonuclease (Haemophilus influenzae)	46	30	357

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match (nt)	Match (nt)	Match gene name	A sim	% ident	length (nt)
165	1	1	402	gi 450688		hadM gene of Ecopri gene product [Escherichia coli] pir S38437 S38437 hadM protein - Escherichia coli pir S09629 S09629 hypothetical protein A - Escherichia coli (SUB 60-520)	46	33	402
4176	1	673	338	gi 551460		FIM-C 1 gene product [Xenopus laevis]	46	31	336
37	7	4813	5922	gi 606064		ORF_408 [Escherichia coli]	45	24	1110
38	16	11699	112004	gi 452152		protein tyrosine phosphatase (PTP-BAS, type 2) [Homo sapiens]	45	24	106
87	2	1748	2407	gi 1064813		homologous to sp:PHOK_BACSU [Bacillus subtilis]	45	23	663
103	12	14182	13385	gi 1001307		hypothetical protein [Synchocystis sp.]	45	22	798
117	14	14791	13181	gi 1204389		H influenzae predicted coding region H10131 [Haemophilus influenzae]	45	23	981
145	4	4483	3461	gi 320578		open reading frame [Mus musculus]	45	20	1023
170	6	6329	4965	gi 238657		AppC-cytochrome d oxidase, subunit I homolog [Escherichia coli, K12, eptide, 514 aa]	45	27	1365
206	2	5230	4346	gi 1222056		aminotransferase [Haemophilus influenzae]	45	27	885
228	1	60	716	gi 160299		glutamic acid-rich protein [Plasmodium falciparum] pir A56514 A56514 glutamic acid-rich protein precursor - Plasmodium falciparum	45	23	657
248	1	2	1015	gi 1255425		C3308 2 gene product [Caenorhabditis elegans]	45	23	1018
313	3	4339	3128	gi 581140		NADH dehydrogenase [Escherichia coli]	45	30	1212
332	1	914	459	gi 870966		[F47A4.2 [Caenorhabditis elegans]	45	20	456
344	1	3	221	gi 171225		kinesin-related protein [Saccharomyces cerevisiae]	45	26	219
441	2	1501	1073	gi 162863		replication initiation protein [Bacillus subtilis] pir B26580 B26580 replication initiation protein - Bacillus subtilis	45	27	459
672	1	2	982	gi 1511334		M. jannaschii predicted coding region MJ1323 [Methanococcus jannaschii]	45	22	981
761	3	1315	851	gi 606180		ORF_510 [Escherichia coli]	45	24	495
886	3	379	846	gi 726426		similar to protein kinases and C. elegans proteins F37C12.8 and 37C12.5 [Caenorhabditis elegans]	45	30	468
948	1	3	473	gi 156400		myosin heavy chain (isozyme unc-54) [Caenorhabditis elegans] pir A33958 MKW myosin heavy chain B - Caenorhabditis elegans sp P02566 MYSB-CABEL MYOSIN HEAVY CHAIN B (MHC B)	45	25	471
1158	1	2	376	gi 341155		transmission-blocking target antigen [Plasmodium falciparum]	45	35	375
2551	1	4	285	gi 1276705		ORF287 gene product [Porphyra purpurea]	45	28	282
3967	1	42	374	gi 976025		HraA [Escherichia coli]	45	28	333

TABLE 2



S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
52	7	6931	5846	gi 467378	unknown [Bacillus subtilis]	44	22	1086
138	8	6475	6849	gi 173028	thioredoxin II [Saccharomyces cerevisiae]	44	28	375
221	5	7032	5617	gi 153490	tetracenomycin C resistance and export protein [Streptomyces laevis]	44	21	1416
252	2	1331	1122	gi 1204989	hypothetical protein (GB U00022.9) [Haemophilus influenzae]	44	30	210
263	2	3265	2093	gi 1136221	carboxypeptidase [Sulfolobus solfataricus]	44	26	1173
365	4	4963	3524	gi 296822	orf1 gene product [Lactobacillus helveticus]	44	31	1440
563	3	1115	1833	gi 1063250	low homology to P20 protein of Bacillus licheniformis and bleomycin acetyltransferase of Streptomyces verticillius [Bacillus subtilis]	44	24	519
544	4	3942	4892	gi 1951460	FIN-C.1 gene product [Xenopus laevis]	44	32	951
792	1	1224	613	gi 205680	high molecular weight neurofilament [Rattus norvegicus]	44	28	612
48	18	11303	11911	gi 1511614	molycoprotein-guanine dinucleotide biosynthesis protein A [Methanococcus jannaschii]	43	27	609
59	8	3665	5128	gi 153490	tetracenomycin C resistance and export protein [Streptomyces laevis]	43	21	1464
59	10	5536	7527	gi 153022	lipase [Staphylococcus epidermidis]	43	22	1992
99	1	1346	681	gi 1419051	unknown [Mycobacterium tuberculosis]	43	21	666
310	8	9402	12134	gi 197526	clumping factor [Staphylococcus aureus]	43	21	2733
412	3	2782	2303	pir A60540 A605	apoptosis surface protein 2 - Plasmodium yoelii (fragment)	43	29	480
519	3	2547	3122	sp Q06510 Q06510	SULFIDE DEHYDROGENASE (FLAVOCYTOCHROME C) FLAVOPROTEIN CHAIN PRECURSOR (Hc 1.8.2-1) (FCSU)	43	23	576
4	113	12053	13321	gi 295671	selected as a weak suppressor of a mutant of the subunit AC10 of DNA dependent RNA polymerase I and III [Saccharomyces cerevisiae]	42	18	1269
94	2	1768	1091	gi 1501027	ORF2 [Trypanosoma brucei]	42	31	678
127	4	5791	4550	gi 42029	ORF1 gene product [Escherichia coli]	42	21	1242
287	3	1515	1036	gi 142790	ORF1, putative [Bacillus firmus]	42	25	480
144	6	4097	1525	gi 140320	ORF 2 (AA 1-203) [Bacillus thuringiensis]	42	30	573
512	3	2167	1115	gi 405957	yefE [Escherichia coli]	42	23	1053
631	3	2434	1223	gi 1580920	rodD (gcaA) polypeptide (AA 1-673) [Bacillus subtilis] pir S06048 S06048 probable rodD protein - Bacillus subtilis sp P3484 P3484 (TAGE-BACSU PROBABLE POLY (GLYCEROL-PHOSPHATE) L-PHA-GLUCOSYLTRANSFERASE (EC 2.4.1.52) (TECHOIC ACID BIOSYNTHESIS PROTEIN E))	42	24	1212

TABLE 2

C. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
685	3	2359	1739	gi11303784	YqoD (Bacillus subtilis)	42	19	621
4132	1	787	395	gi11022810	protein tyrosine phosphatase [Dictyostelium discoideum]	42	25	393
86	2	1375	884	gi1309506	spermidine/spermine N1-acetyltransferase [Mus saxicola] pir[S27026]S27026 spermidine/spermine N1-acetyltransferase - spiny ouse (Mus saxicola)	41	30	492
191	12	114797	114075	gi1124857	lorf4 gene product [Methanosarcina barkeri]	41	22	723
212	6	2150	1127	gi115873	observed 35 kD protein [Mycobacteriophage 15]	41	26	978
213	3	1263	2000	gi1633692	TrsA [Yersinia enterocolitica]	41	18	718
408	4	2625	3386	gi11197634	lorf4: putative transporter: Method: conceptual translation supplied by author [Mycobacterium smegmatis]	41	24	762
542	1	3	1103	gi1457146	thoxy protein [Plasmodium yoelii]	41	21	1101
924	1	2	475	pir1JH01481JH01	nucleolin - rat	41	30	474
1562	1	1	402	gi1552184	asparagine-rich antigen PfA5-2 [Plasmodium falciparum] pir[S27026]S27026 asparagine-rich antigen PfA5-2 - Plasmodium falciparum (fragment)	40	20	402
2395	1	518	261	pir[S4225]S422	hypothetical protein 5 - fowlpox virus	40	18	258
4077	1	1	305	gi1045095	coded for by C. elegans cDNA YK37g1.5; coded for by C. elegans cDNA YK5c9.5; coded for by C. elegans cDNA YK1a9.5; alternatively spliced form of F52C9.8b [Caenorhabditis elegans]	39	21	303
918	1	1003	503	gi11255425	C33C8.2 gene product [Caenorhabditis elegans]	37	25	501
59	12	8294	10636	gi1535260	STAP antigen [Plasmodium reichenowi]	36	24	2343
63	5	3550	8079	gi1298012	EF [Streptococcus suis]	36	19	4510
544	3	2507	3601	gi11015903	ORF YJH15c [Saccharomyces cerevisiae]	35	22	1095
63	4	1949	3574	gi1552195	circumsporozoite protein [Plasmodium falciparum] sp1K05691[CSE_PfAFL CIRCUMSPOROZOITE PROTEIN (CS) (FRAGMENT)]	32	27	1626

TABLE 2

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
4	1	1234	692
4	3	1712	2278
4	4	3703	1032
4	14	13073	12585
5	2	2539	1601
5	3	1532	1731
5	7	4741	4550
5	9	7939	6423
5	12	8711	8547
6	4	2359	1982
8	1	349	176
11	8	5144	5983
11	9	5988	6498
11	10	6472	6284
11	16	10954	11271
12	5	5352	4942
12	6	4596	4862
15	3	1895	1650
16	10	11263	10835
18	2	1093	917
20	9	9125	7764
20	10	8571	8230
20	12	9201	8603
20	13	12158	10470
23	1	674	339
23	6	6138	5485
23	8	6376	5942

TABLE 3

S aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
23	9	7651	6881
23	15	12618	12830
24	4	4556	4185
24	6	5642	5241
25	2	1824	2402
31	2	505	849
31	3	1177	1524
31	4	2454	3005
32	2	765	1388
32	9	7952	8575
32	10	8591	8728
32	11	9738	9379
32	12	10797	10087
34	2	1315	1049
36	7	5226	5801
36	11	7575	7261
36	12	7424	7621
37	4	3158	2984
38	2	1585	980
38	11	6425	6868
38	20	16982	16371
38	26	20253	20804
38	27	20722	21264
39	1	1	627
40	1	605	404
43	1	796	428
44	4	2674	2324

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Accession ID	ORF ID	Start (nt)	Stop (nt)
44	5	2484	3263
44	14	10587	10129
44	20	13724	13536
44	21	13596	13994
45	7	6575	6297
46	8	6165	6520
46	12	10449	10976
46	17	15032	15424
47	1	288	1079
48	9	7620	7778
50	1	1612	962
50	2	1621	1316
51	1	718	370
51	5	2520	2245
53	1	442	287
53	7	6705	6319
54	7	9014	8709
55	1	592	326
55	3	1052	786
56	1	1	261
56	3	1551	1228
56	4	1970	1560
56	17	19092	18712
57	4	3694	3521
57	8	5436	5822
58	9	8885	8553
59	3	1366	1509

TABLE 3

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55

Sequences - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
59	6	3026	2802
59	7	3770	3570
59	9	4946	4563
59	11	7518	8378
59	13	10401	16403
62	2	2696	1521
62	11	5440	5757
63	1	1	336
67	1	900	1781
67	2	1774	2610
67	3	2591	3904
67	8	7110	6955
68	1	78	326
70	6	6761	5199
70	11	8935	8645
77	3	1590	1192
79	2	1509	1228
79	3	1411	1791
81	1	2	403
85	9	8300	8653
85	10	8969	8781
86	3	1426	1232
87	8	9187	9366
88	3	1620	1922
89	1	3	161
89	7	5042	4878
91	1	1098	550

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
91	3	1938	3141
92	2	449	928
92	3	1958	1467
92	9	5638	6024
94	1	661	332
94	3	2445	1813
94	4	2583	2197
96	11	110601	11050
99	6	4672	4523
99	7	5014	4784
100	8	7658	7287
102	7	4697	4368
103	3	2496	2035
104	1	2	694
104	2	699	1277
105	1	1235	693
105	3	3233	2655
105	1	3	221
106	3	1209	1355
107	1	1081	542
109	4	4025	3651
109	13	11625	11996
109	14	11981	12268
109	20	117401	11688
110	1	2	760
114	10	8764	9384
116	1	1	309

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
116	3	6273	4462
116	8	11049	9976
116	9	10313	10158
120	5	7703	3320
120	6	4270	3869
120	11	9290	9844
121	2	417	569
126	3	1090	818
127	3	2648	3196
127	5	4084	4395
131	6	6773	6438
132	2	715	1695
134	1	2	667
135	2	512	258
135	3	1124	729
138	1	3	152
138	7	6008	6463
140	1	2040	1012
140	2	2059	1513
140	5	2187	2743
142	2	1380	2368
142	7	8830	7586
143	7	7290	6502
146	1	1227	640
146	1	2	511
146	3	502	1350
146	4	3673	2540



TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
146	5	2874	3071
147	1	1	339
149	11	3956	3615
149	12	4036	3785
149	13	4507	4145
149	15	4807	4610
149	16	5495	5049
149	18	5739	5491
149	21	7416	7054
149	23	9216	8521
149	24	9681	9106
149	25	10679	9897
150	2	2303	1587
154	3	1795	1508
154	8	6586	6398
154	14	12704	12147
154	15	13531	12803
156	1	315	593
157	3	1183	2232
158	2	1471	1064
159	3	452	808
161	2	876	1808
161	6	4653	4279
161	7	4803	4540
161	8	4886	4717
161	11	5817	5638
163	2	1604	840

TABLE 3

*S. aureus* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
163	5	2796	2344
163	7	2952	2647
163	9	4905	5132
164	3	1338	1147
166	3	5213	4854
168	4	2500	2888
168	5	3595	4158
170	3	2517	2777
171	2	2277	1450
171	11	12576	11125
172	1	3	278
172	2	1940	1149
173	1	1289	708
173	5	7001	6114
174	2	593	1105
175	3	2552	2890
175	5	3820	3335
175	7	4342	4506
182	4	5477	4986
184	5	6043	5702
188	2	1210	1755
188	4	2647	2994
189	6	2614	3039
190	3	1998	2564
191	1	1	153
191	2	950	669
191	10	11786	11039

S. aureus - putative coding regions of novel proteins not similar to known proteins

Contig ID	OHF ID	Start (nt)	Stop (nt)
191	11	12902	13263
192	1	91	426
195	3	2306	1932
195	5	2899	2606
198	2	1016	1591
201	1	170	625
203	2	783	1466
206	6	8930	7815
206	12	113947	11616
206	21	128208	127960
212	2	170	817
212	3	796	1167
212	7	3178	3436
212	9	3749	4075
213	1	1	705
214	2	1076	570
214	6	4064	3738
214	9	6600	6995
214	10	7864	7469
217	1	1927	965
218	1	178	657
218	3	1776	2156
220	2	1851	1369
220	3	3251	2262
220	7	8275	7208
220	8	10244	8661
220	9	11796	10216

TABLE 3

TABLE 3

*S. aureus* - Putative coding regions of novel proteins not similar to known proteins

Contig	ORF	Start	Stop
ID	ID	(nt)	(nt)
221	4	1095	2613
221	9	11428	10757
226	1	3	659
226	2	2196	1459
226	3	1476	1961
227	1	2	487
227	2	460	975
227	4	1855	2121
227	5	2052	2345
227	6	4760	3768
227	9	5591	6367
228	5	2503	2877
228	6	2846	3526
233	7	3944	3762
236	2	809	579
238	2	1975	1391
239	2	1417	905
241	5	4495	4334
242	2	1677	1363
243	1	127	576
244	1	1291	647
244	2	3035	1962
245	2	1614	1258
246	1	69	215
246	4	738	1733
249	3	3906	3712
250	1	494	249

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
254	1	1	156
256	2	956	1144
257	3	3700	3227
260	4	4906	4580
261	4	2196	2606
261	6	3218	3681
264	2	155	439
264	5	5252	4533
264	6	4239	5107
267	2	1323	931
268	4	5140	4700
272	1	862	446
272	3	1200	1439
272	9	4691	4909
272	10	6469	6035
276	6	1746	1901
278	1	224	553
278	5	3299	3448
278	7	4849	5127
285	2	551	736
288	3	1756	1950
288	5	2055	2276
289	1	2107	1055
290	2	2234	1932
291	2	332	622
291	5	1545	2051
295	3	1606	1349

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
295	4	2728	2141
295	5	2220	2762
297	2	788	465
298	1	2	205
300	2	2380	1928
301	7	2794	2624
304	1	3	194
306	1	109	654
306	5	4016	4257
307	1	674	339
307	8	3645	3995
308	1	1	654
308	2	1120	599
308	4	7643	2332
313	2	2314	1919
314	1	10	702
316	2	982	1341
316	6	2758	3165
317	1	2	1114
317	3	4570	3458
321	6	5645	5217
321	7	6319	6140
321	8	7450	6794
322	2	827	543
326	2	165	1112
326	3	1117	1467
328	1	936	469

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID (nt)	ORF ID (nt)	Start (nt)	Stop (nt)
328	5	3452	3276
329	1	3	719
329	2	781	1212
329	3	1471	1813
330	1	576	289
330	3	1447	1623
332	3	2353	2204
332	7	4971	5138
333	2	2295	3128
335	1	864	433
337	2	95	526
340	2	1658	1356
341	1	3	281
341	3	2476	3192
341	5	3618	3844
341	6	3929	4558
344	5	3197	2889
345	1	1532	768
346	2	221	592
350	3	1410	1598
352	2	2178	1765
352	3	7116	4596
352	7	7967	8404
352	8	8906	9247
352	9	10171	9854
359	1	1	546
362	1	3	656

TABLE 3

S aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
364	2	2158	1808
364	8	10774	10714
365	2	1612	1313
365	5	4680	4090
365	7	4980	6239
366	3	520	1719
367	3	906	1085
368	1	748	494
375	1	2	136
380	3	1351	1097
389	1	1	276
390	1	2	877
390	2	1373	1549
391	2	751	560
395	1	391	197
396	1	2132	1068
398	3	1344	1141
399	1	176	669
401	3	566	847
402	2	100	465
404	8	5561	5170
408	2	3507	2269
408	3	2875	2672
408	5	3524	4423
410	3	2111	1890
413	1	880	488
416	1	607	320



TABLE 3

5 *Salmonella* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
416	2	578	847
416	3	2195	1590
417	1	3	179
417	2	161	616
420	2	788	513
422	2	357	677
431	2	856	1407
432	2	446	1084
433	1	1	417
433	3	2311	2033
434	1	942	535
434	2	2089	1235
440	1	1	450
442	2	1269	3320
443	3	1873	1520
444	1	1	686
444	7	6761	6366
451	1	940	614
453	2	896	636
453	8	3833	4786
453	9	4718	4512
453	10	4937	4731
455	1	434	219
455	2	472	930
459	1	265	687
462	1	2	247
466	2	1494	907

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig	ORF	Start	Stop
10	(nt)	(nt)	(nt)
467	1	654	349
468	1	2	250
469	1	1488	925
469	3	2386	3372
469	4	3464	3706
470	1	77	538
470	6	4098	3694
470	7	6310	5685
470	9	7351	8181
470	10	8175	9733
471	1	940	500
471	2	1562	1017
476	1	70	267
477	1	2	760
477	3	3764	2081
477	4	2066	2332
480	5	4016	4261
481	2	956	480
486	3	613	774
487	6	3795	2112
488	1	715	359
492	1	127	675
493	1	2	520
493	2	496	1242
502	3	1149	1571
504	1	690	346
505	5	4586	4150

TABLE 3

S. aureus - putative ending regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
511	2	1741	1232
512	2	583	747
515	1	639	812
517	4	2179	2511
520	4	2097	2360
520	6	1908	3669
527	1	1	498
528	1	637	315
529	2	1679	1104
530	7	5298	5534
536	1	308	156
538	1	1362	736
538	3	2203	2880
538	5	3531	3121
538	6	4348	3721
540	1	996	664
540	2	1495	1031
541	1	89	433
541	2	719	432
542	2	1048	1272
545	2	1012	734
551	1	2145	1129
555	2	892	704
558	3	1357	1154
558	4	1760	1458
558	5	2105	1821
558	6	2166	2020

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
558	7	2536	2522
558	8	3053	2802
558	9	3986	3453
580	1	475	921
585	3	1706	1485
571	1	308	156
571	3	994	1206
577	1	2	199
577	2	163	453
579	1	1	477
579	2	1784	1200
583	1	1988	996
585	1	946	539
587	1	22	573
588	2	1896	1372
588	3	1782	1554
590	1	47	334
592	2	1455	1341
593	1	2	775
593	2	817	1122
595	1	87	890
596	3	1593	1415
602	1	8	169
603	5	1071	1469
606	1	322	768
607	5	1444	1226
610	1	1029	541

TABLE 3

S. aureus - Relative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
612	1	3	500
616	1	991	650
617	2	736	491
622	1	36	347
625	4	2046	2549
627	1	67	210
628	1	901	452
631	3	4789	4004
634	1	1448	759
636	1	189	368
636	2	1929	1063
637	2	2323	1994
638	1	227	1081
639	1	518	261
639	2	3377	811
641	1	118	444
642	3	1615	1331
642	4	2260	1847
643	1	3	608
645	4	1534	1758
645	6	2025	2321
645	7	2940	2488
648	1	2	1045
680	1	77	601
680	2	576	872
681	1	1725	961
684	2	89	304

5. Accession - Putative coding regions of novel proteins not similar to known proteins

Accession	ORF	Start	Stop
no	id	(nt)	(nt)
667	1	3	413
668	1	1	330
671	2	812	516
673	1	3	338
674	2	865	584
679	1	1	237
679	3	1589	1906
688	1	1236	835
688	2	1352	1077
694	1	3	143
696	2	818	432
706	1	26	224
709	3	1143	1449
711	1	3	908
715	1	3	167
716	1	2	637
721	1	133	570
722	1	763	383
723	1	1656	829
723	2	1498	1112
727	1	2	472
729	1	268	441
731	1	130	828
735	1	2	214
736	1	3	782
738	1	2	288
742	1	3	230

TABLE 3

TABLE 3

S. aureus - putative coding regions of novel proteins not similar to known proteins

Coding ID	ORF (nt)	Start (nt)	Stop (nt)
745	3	1:48	780
748	2	282	464
749	1	685	344
751	1	901	452
755	1	97	522
755	2	520	918
758	2	663	400
764	2	1033	746
767	1	1	405
768	1	2	373
771	1	1058	534
778	1	1735	902
785	1	1790	1023
787	1	1260	631
791	1	3	224
799	1	15	260
804	1	304	711
805	1	3	680
808	1	219	842
810	1	2221	1112
810	2	1774	1442
812	1	38	979
817	1	714	358
818	2	487	1104
819	2	1529	1012
819	3	1748	1419
820	1	195	1064

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
828	1	506	255
829	1	48	800
830	1	578	291
832	1	594	298
835	1	320	796
840	3	491	709
845	1	912	457
850	2	303	449
853	1	715	359
860	1	2	256
864	1	18	410
864	2	383	715
864	6	1676	1818
870	1	1	588
873	1	906	454
875	1	584	294
877	1	1661	1020
878	1	981	544
879	1	1567	785
881	1	1	243
882	1	389	604
890	1	2	508
905	1	793	358
906	1	852	544
912	1	373	188
913	1	3	290
913	2	1092	547



TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig	ORF	Start	Stop
1B	10	(nt)	(nt)
915	1	6	161
915	2	169	402
921	1	126	386
927	1	1578	808
928	1	2	385
929	1	2	400
932	1	2	400
934	1	1	384
936	1	1052	528
937	1	2	616
945	1	220	645
945	2	649	1242
946	1	1702	950
949	1	1	270
951	1	3	362
955	1	3	143
960	1	723	400
963	1	1	162
965	1	690	346
966	1	1079	606
969	1	3	302
971	1	12	170
974	1	319	161
976	1	692	348
977	1	2	211
982	1	1926	982
984	1	589	296

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig	ORF	Start	Stop
10	10	(nt)	(nt)
987	1	3	487
993	1	1	525
994	1	920	549
1006	1	557	318
1016	1	624	313
1015	1	2	463
1016	1	288	345
1019	1	1205	660
1022	1	839	474
1024	1	595	299
1024	2	276	431
1030	1	673	338
1032	1	355	179
1040	1	774	399
1043	1	3	269
1044	2	115	399
1047	1	1	159
1051	1	704	354
1051	2	1233	733
1063	1	2	400
1069	1	2	148
1069	2	769	533
1075	1	707	399
1077	1	97	405
1081	1	58	438
1086	1	1	384
1087	2	246	431

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig (ORF)	Start (nt)	Stop (nt)
1088	1	374
1096	1	474
1098	1	1015
1100	1	1020
1100	2	1520
1101	1	703
1102	1	385
1107	1	2
1114	1	3
1115	1	2
1119	1	22
1129	1	40
1132	1	360
1133	1	609
1144	1	446
1147	1	558
1153	1	1
1154	1	3
1159	1	1
1161	1	341
1184	1	427
1171	1	19
1171	2	108
1183	1	2
1195	1	355
1196	1	1
1200	1	33

TABLE 3

S. aureus Putative coding regions of novel proteins not similar to known proteins

Config	ORF	Start	Stop
ID	ID	(nt)	(nt)
1203	2	129	464
1222	2	105	401
1232	1	1	387
1240	1	2	175
1247	1	520	311
1271	1	412	221
1286	1	2	555
1295	1	1	165
1306	1	167	185
1314	2	158	631
1316	1	58	570
1339	1	184	193
1370	1	1	402
1371	1	1	345
1374	1	710	357
1378	1	2	400
1392	1	3	413
1411	1	202	432
1433	1	311	167
1450	1	2	256
1453	1	295	149
1471	1	721	398
1477	1	869	639
1502	1	794	399
1518	1	126	449
1534	1	283	143
1546	1	3	401

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
1547	1	506	255
1583	1	3	350
1587	1	3	563
1602	2	170	679
1629	1	1	402
1665	1	468	235
1760	1	625	314
1762	1	3	200
1876	2	119	286
1895	1	2	379
1931	1	798	400
1976	2	715	383
2055	2	252	401
2056	1	331	167
2150	1	523	263
2157	1	794	399
2164	1	564	283
2175	1	218	400
2212	1	492	331
2338	1	732	367
2342	1	3	167
2352	1	330	166
2352	2	622	398
2355	1	47	352
2356	1	679	341
2359	1	301	152
2421	1	296	150

TABLE 3

S. aureus - putative coding regions of novel proteins not similar to known proteins

Contig	ORF	Start	Stop
ID	ID	(nt)	(nt)
3046	1	367	165
3049	1	553	278
3050	1	3	314
3052	1	504	253
3065	1	2	157
3070	1	357	190
3075	1	440	222
3080	1	1	285
3092	1	120	162
3093	1	411	750
3100	1	52	237
3103	1	47	298
3118	1	344	174
3123	1	2	145
3127	1	1	147
3138	1	316	169
3142	1	388	203
3144	1	664	386
3151	1	137	170
3155	2	202	384
3168	1	12	176
3205	1	288	145
3282	1	1	150
3303	2	239	400
3371	2	211	399
3558	1	2	148
3558	2	36	401

TABLE 3

*S. aureus* - putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
3046	1	367	185
3049	1	553	278
3050	1	3	314
3052	1	504	253
3065	1	2	157
3070	1	357	190
3075	1	440	222
3080	1	1	285
3092	1	320	162
3093	1	411	250
3100	1	52	237
3103	1	47	298
3116	1	344	174
3123	1	2	145
3127	1	1	147
3138	1	336	169
3142	1	388	203
3144	1	664	386
3151	1	337	170
3155	2	202	384
3168	1	12	176
3205	1	288	145
3282	1	1	150
3303	2	239	400
3371	2	211	399
3558	1	2	148
3558	2	36	401

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
3568	1	751	377
3595	1	757	380
3618	1	2	238
3618	2	130	402
3642	1	86	358
3622	2	664	398
3642	1	876	439
3649	1	781	398
3651	1	625	314
3664	1	487	637
3674	1	55	402
3677	1	619	311
3704	1	1	402
3726	1	535	269
3765	1	510	256
3779	1	554	357
3794	1	266	135
3794	2	667	377
3796	2	638	375
3801	1	474	262
3806	1	453	298
3807	1	42	389
3815	1	798	400
3827	1	3	320
3842	1	781	392
3853	1	671	399
3855	1	1	324



TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
3857	1	2	225
3861	1	590	297
3865	1	695	399
3897	1	3	173
3897	2	143	400
3898	2	225	401
3921	2	103	342
3927	1	70	175
3930	1	76	234
3946	2	651	382
3951	2	105	377
3965	1	646	344
3973	1	795	400
3981	1	3	311
3998	1	3	356
4001	1	481	296
4003	1	90	335
4018	1	2	259
4018	2	186	401
4021	1	1	345
4043	1	3	344
4054	1	3	344
4066	1	1	150
4070	1	1	324
4072	2	187	390
4073	1	1	285
4077	2	127	372

TABLE 3

*S. aureus* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
4083	1	3	359
4090	1	27	388
4101	1	103	297
4105	1	1	306
4107	1	570	286
4119	1	629	339
4121	1	740	372
4123	1	3	230
4127	1	3	341
4128	1	2	331
4130	1	768	415
4146	1	97	381
4157	1	3	206
4186	1	505	254
4224	1	510	256
4239	1	1	348
4242	1	709	356
4252	1	589	286
4253	1	1	174
4256	1	568	323
4258	2	498	334
4267	1	284	144
4271	1	2	304
4287	1	303	163
4289	1	471	319
4302	1	153	305
4304	1	1	186

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
4304	2	96	314
4306	1	2	151
4318	1	576	289
4322	1	5	148
4331	1	439	221
4331	2	528	364
4338	1	728	399
4346	1	471	277
4367	2	117	311
4373	1	2	288
4381	1	574	326
4384	1	614	309
4397	1	9	311
4402	1	1	249
4403	1	606	328
4406	1	1	317
4411	1	2	280
4411	2	697	398
4412	1	2	366
4418	1	3	230
4424	1	601	398
4443	1	427	215
4471	1	643	323
4478	1	540	271
4482	1	50	289
4489	1	601	302
4491	1	12	206

TABLE 3

*S. aureus* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
4495	1	3	179
4496	1	500	252
4500	1	130	306
4511	1	493	248
4518	1	1	246
4526	1	480	241
4527	1	2	163
4532	1	3	239
4542	1	11	175
4567	1	36	200
4573	1	1	231
4578	1	642	322
4619	1	1	180
4620	1	349	176
4662	1	1	246
4669	1	2	157
4680	1	28	183
4690	1	344	174

Table 4

ORF	SEQ ID NO	BLAST	Antigenic Regions			
		HOMOLOG	Region 1	Region 2	Region 3	Region 4
168_6	5192	lipoprotein	36-45	84-103	152-161	176-185
238_1	5193	chrA	21-39	48-58	84-95	232-249
51_2	5194	OppB gene product (B. sub)	20-36	70-79	100-112	121-131
278_3	5195	lipoprotein 1	20-29	59-73	85-97	162-171
276_2	5196	lipoprotein	21-33	65-74	177-186	211-220
45_4	5197	ProX	28-37	59-69	85-100	120-129
315_8	5198	hypothetical protein	45-54	88-97	182-192	243-253
154_15	5199	unknown	31-40	48-58	79-88	95-104
228_3	5200	unknown	25-38	40-52	64-74	80-89
228_6	5201	unknown	29-41	89-101	128-143	173-184
50_1	5202	unknown	21-33	52-61	168-182	197-206
112_7	5203	iron-binding periplasmic	21-31	58-67	92-101	111-120
442_1	5204	unknown	30-39	91-100	122-137	182-192
66_2	5205	unknown	50-59	104-116	127-136	167-182
304_2	5206	Q-binding periplasmic	19-28	48-57	75-84	103-116
44_1	5207	hypothetical protein	27-36	86-95	129-138	192-201
161_4	5208	SphX	27-44	149-161	166-175	201-210
46_5	5209	cmpC (permease)	21-33	61-70	83-92	100-109
942_1	5210	traH [Plasmid pSK41]	83-92	109-118	127-142	
5_4	5211	ORF (S. aureus)	12-22	87-96	111-120	151-160
20_4	5212	peptidoglycan hydrolase (S.	24-34	129-138	141-150	161-171
328_2	5213	lipoprotein (H. flu)	81-90	123-133	290-299	
520_2	5214	fibronectin binding protein	44-54	63-79	81-90	95-110
771_1	5215	emm1 gene product (S. pyi	30-39	65-82	96-106	112-121
999_1	5216	predicted trithorax prot. (D	7-16	120-129	157-166	
853_1	5217	ORF2136 (Marchantia polyr	43-52	88-97	102-111	
287_1	5218	psaA homolog	13-22	28-44	72-82	114-124
288_2	5219	cell wall enzyme	14-23	89-98		
596_2	5220	penicillin binding protein 2b	40-49	59-68	76-87	106-115
217_5	5221	fibronectin/fibrinogen bindi	28-37	40-49	62-71	93-111
217_6	5222	fibronectin/fibrinogen bp	10-19	31-40	54-62	73-92
528_3	5223	myosin cross reactive prote	4-13	29-47	60-73	90-99
171_11	5224	EF	20-31	91-110		
63_4	5225	penicillin binding protein 2b	12-21	59-68	95-104	
353_2	5226		46-55	62-71		
743_1	5227	29 kDa protein in fimA regi	23-32	68-79	94-103	175-184
342_4	5228	Twitching motility	10-19	48-60	83-92	111-121
69_3	5229	arabinogalactan protein	97-106	132-141	158-167	180-189
70_6	5230	nodulin	36-45	48-57	137-160	179-188
129_2	5231	glycerol diester phosphodie	8-17	41-50	55-74	97-106
58_5	5232	PBP (S. aureus)	26-35	70-79	117-126	152-161
188_3	5233	MHC class II analog (S. aure	72-81	94-103	115-124	136-145
236_6	5234	histidine kinase domain (Dic	24-33	52-67	81-94	106-121
310_8	5235	clumping factor (S. aureus)	59-71	77-86	93-102	118-127
601_1	5236	novel antigen/ORF2 (S. au	45-54	91-104	108-117	186-195
544_3	5237	ORF YJR151c (S. cerevisae	76-90	101-111	131-140	154-164
662_1	5238	MHC class II analog (S. aure	22-32	71-80	89-98	114-122
87_7	5239	5' nucleotidase precursor ('	29-45	62-71	105-114	125-137
120_1	5240	B65G gene product (B. sub	102-111			

Table 4

5	ORF	Antigenic Regions (cont)				
	Region 5	Region 6	Region 7	Region 8	Region 9	Region 10
	168_6	244-272	303-315			
	238_1	260-269	291-301	308-317		
	51_2	140-152	188-208	211-220	256-266	273-283
10	278_3	198-209				
	276_2	255-268				
	45_4	177-199	221-230	234-243	268-279	284-293
	316_8					304-313
	154_15	148-157	177-187	202-211		
15	228_3	101-119	139-154	166-181		
	228_6					
	50_1					
	112_7	136-149	197-211	218-229	253-273	
	442_1	199-210	247-257	264-277	287-309	
20	66_2					
	304_2	178-187	250-259			
	44_1					
	161_4					
	46_5	131-141	162-176	206-215	243-252	264-273
25	942_1					285-294
	5_4	189-205	230-239	246-264	301-318	340-354
	20_4	202-212	217-234	260-275	314-336	366-373
	328_2					378-387
	520_2					380-391
30	771_1	145-154				
	999_1					
	853_1					
	287_1	154-164				
	288_2					
35	596_2	121-130				
	217_5	244-253	259-268	288-297	302-311	
	217_6	144-158	174-183	188-197	207-216	226-242
	528_3					
40	171_11					
	63_4					
	353_2					
	743_1	197-207				
	342_4					
45	69_3	195-211				
	70_6	206-215	263-272	291-301	331-340	358-371
	129_2	117-127	141-157	168-183	202-211	222-231
	58_5	184-203	260-269	275-299	330-344	372-381
	188_3					424-433
50	236_6	138-147	163-172	187-198	244-261	268-278
	310_8	131-140	144-153	177-186	190-199	204-213
	601_1	208-218				308-317
	544_3	170-179	184-193	224-235	274-287	327-336
	662_1					352-361
55	87_7					
	120_1					

Table 4

	ORF	Antigenic Regions (cont)					
		Region 11	Region 12	Region 13	Region 14	Region 15	Region 16
5	168_6						
	238_1						
	51_2						
10	278_3						
	276_2						
	45_4						
	316_8						
	154_15						
15	228_3						
	228_6						
	50_1						
	112_7						
	442_1						
20	66_2						
	304_2						
	44_1						
	161_4						
	46_5	306-315					
25	942_1						
	5_4	393-407	416-426	456-465			
	20_4	396-405	410-419	461-481			
	328_2						
	520_2						
30	771_1						
	999_1						
	853_1						
	287_1						
	288_2						
35	596_2						
	217_5						
	217_6						
	528_3						
	171_11						
40	63_4						
	353_2						
	743_1						
	342_4						
	69_3						
45	70_6	453-471	506-515				
	129_2	296-315					
	58_5						
	188_3						
50	236_6	358-377	410-423	428-439	442-457	467-476	480-493
	310_8	238-251	256-275	281-290	296-310	314-333	338-347
	601_1						
	544_3						
	662_1						
55	87_7						
	120_1						

Table 4

	ORF	Antigenic Regions (cont)					
		Region 17	Region 18	Region 19	Region 20	Region 21	Region 22
5	168_6						
	238_1						
	51_2						
10	278_3						
	276_2						
	45_4						
	316_8						
	154_15						
15	228_3						
	228_6						
	50_1						
	112_7						
	442_1						
20	66_2						
	304_2						
	44_1						
	161_4						
	46_5						
25	942_1						
	5_4						
	20_4						
	328_2						
	520_2						
30	771_1						
	999_1						
	853_1						
	287_1						
	288_2						
35	596_2						
	217_5						
	217_6						
	528_3						
	171_11						
40	63_4						
	353_2						
	743_1						
	342_4						
	69_3						
45	70_6						
	129_2						
	58_5						
	188_3						
	236_6						
50	310_8	357-366	370-379	429-438	443-452	478-487	551-560
	601_1						
	544_3						
	662_1						
55	87_7						
	120_1						



Table 4

5	ORF	Antigenic Regions (cont)					Region 28
		Region 23	Region 24	Region 25	Region 26	Region 27	
	168_6						
	238_1						
	51_2						
10	278_3						
	276_2						
	45_4						
	316_8						
	154_15						
15	228_3						
	228_6						
	50_1						
	112_7						
	442_1						
20	66_2						
	304_2						
	44_1						
	161_4						
	46_5						
25	942_1						
	5_1						
	20_4						
	328_2						
	520_2						
30	771_1						
	999_1						
	853_1						
	287_1						
	288_2						
35	596_2						
	217_5						
	217_6						
	528_3						
	171_11						
40	63_4						
	353_2						
	743_1						
	342_4						
	69_3						
45	70_6						
	129_2						
	58_5						
	188_3						
	236_6						
50	310_8	622-632	670-685	708-718	823-836	858-867	877-886
	601_1						
	544_3						
	662_1						
	87_7						
55	120_1						

Table 4

	ORF	Antigenic Regions (cont)	
		Region 29	Region 30
5	168_6		
	238_1		
	51_2		
10	278_3		
	276_2		
	45_4		
	316_8		
	154_15		
15	228_3		
	228_6		
	50_1		
	112_7		
	442_1		
20	66_2		
	304_2		
	44_1		
	161_4		
	46_5		
25	942_1		
	5_4		
	20_4		
	328_2		
	520_2		
30	771_1		
	999_1		
	853_1		
	287_1		
	288_2		
35	596_2		
	217_5		
	217_6		
	528_3		
	171_11		
40	63_4		
	353_2		
	743_1		
	342_4		
	69_3		
45	70_6		
	129_2		
	58_5		
	188_3		
	236_6		
50	310_8		
	601_1		
	544_3		
	662_1		
55	87_7		
	120_1		

Table 4

ORF		BLAST	Antigenic Regions			
		HOMOLOG	Region 1	Region 2	Region 3	Region 4
46_1	5241	aldehyde dehydrogenase	8-17	36-52	83-96	112-121
63_4	5242	glycerol ester hydrolase (P	9-26	57-73	93-107	123-133
174_6	5243	ketopantoate hydroxymeth	71-80	203-212	242-254	265-274
206_16	5244	ornithine acetyltransferase	1-10	34-43	54-63	194-210
267_1	5245	NaH-antiporter protein (E. f	120-129	332-347	398-408	
322_1	5246	acriflavin resistance protein	58-75	153-164	203-231	264-284
415_2	5247	transport ATP-binding prote	108-126	218-227	298-308	315-334
214_3	5248	2-nitropropane dioxygenase	123-136	216-233	283-292	297-306
587_3	5249	clumping factor	5-14	43-54	59-68	76-95
685_1	5250	signal peptidase	59-68	72-81	86-95	99-108
54_3	5251	fibronectin binding protein I	23-32	37-46	50-59	89-98
54_4	5252	fibronectin binding protein I	43-52	66-75	95-104	147-156
54_5	5253	fibronectin binding protein I	49-60	81-90		
54_6	5254	fibronectin binding protein I	55-71	82-97	139-158	175-186
328_1	5255	lipoprotein (H. flu)	11-20	61-70	96-105	

Table 4

ORF	Antigenic Regions (cont)					
	Region 5	Region 6	Region 7	Region 8	Region 9	Region 10
46_1	215-242	333-352	376-385	416-432	471-487	
63_4	145-154	191-202	212-223	245-265	274-283	291-300
174_6						
206_16	239-259	275-284				
267_1						
322_1	298-319	350-359				
415_2	344-353	371-380	395-404	456-465	486-495	518-527
214_3	318-337	365-375				
587_3	106-115	142-151	156-166	173-182	186-198	204-213
685_1	113-122	130-145				
54_3	128-138	185-194	217-226	251-260	268-277	295-305
54_4	175-188	191-200	203-212	220-229		
54_5						
54_6	220-230	287-304	317-326	344-353	364-373	378-387
328_1						

Table 4

ORF	Antigenic Regions			(cont)		
	Region 11	Region 12	Region 13	Region 14	Region 15	Region 17
46_1						
63_4	306-315	319-328	366-376	395-420	453-462	467-476
174_6						
206_16						
267_1						
322_1						
415_2	539-555					
214_3						
587_3	217-226	278-287	318-327	332-342	351-360	377-386
685_1						
54_3	316-325	329-345	355-372	387-396	416-425	438-448
54_4						
54_5						
54_6	396-407	427-436	514-531	541-550	569-578	612-622
328_1						

Table 4

ORF	Antigenic Regions			(cont)		
	Region 18	Region 19	Region 20	Region 21	Region 22	Region 23
46_1						
63_4	485-500	513-525				
174_6						
206_16						
267_1						
322_1						
415_2						
214_3						
587_3	396-405	426-442	459-470	485-494	505-514	531-562
685_1						
54_3	455-462	472-491	517-536			
54_4						
54_5						
54_6	639-648	673-681	703-715	723-732	749-760	772-788
328_1						

Table 4

ORF	Antigenic Regions (cont)					
	Region 24	Region 25	Region 26	Region 27	Region 28	Region 29
46_1						
63_4						
174_6						
206_16						
267_1						
322_1						
415_2						
214_3						
587_3	567-578	584-601	607-840	844-854	858-870	877-886
685_1						
54_3						
54_4						
54_5						
54_6	793-802	811-826	834-848	866-876	893-903	907-918
328_1						

Table 4

ORF	Antigenic Regions (cont)	
	Region 30	Region 31
46_1		
63_4		
174_6		
206_16		
267_1		
322_1		
415_2		
214_3		
587_3	889-911	927-936
685_1		
54_3		
54_4		
54_5		
54_6	925-944	951-997
328_1		

SEQUENCE LISTING

5

(1) GENERAL INFORMATION:

(i) APPLICANT:

10

(A) NAME: Human Genome Sciences, Inc.

(B) STREET: 9410 Key West Avenue

(C) CITY: Rockville

(D) STATE: Maryland

15

(E) COUNTRY: US

(F) POSTAL CODE: 20850

(ii) TITLE OF INVENTION: Staphylococcus aureus Poly-  
nucleotides and Sequences

20

(iii) NUMBER OF SEQUENCES: 5255

(v) COMPUTER READABLE FORM:

25

(A) MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage

(B) COMPUTER: HP Vectra 486/33

(C) OPERATING SYSTEM: MSDOS version 6.2

(D) SOFTWARE: ASCII Text

30

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:

(B) FILING DATE:

35

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 60/009,861

(B) FILING DATE: 05-JAN-1996

40

(2) INFORMATION FOR SEQ ID NO:1:

45

50

55

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 5895 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

10 TCCATTATGA AGTCACAAGT ACTATAAGCT GCGATGTTAC CAATGTTTTT TAAAAATCCCA 60  
 GTAATAAAAT CAAAAAATAA GTTAAATAAT GTATTCATTT TAAGTCCTCC TTAATAAAGa 120  
 15 aaataGGTAA TAATGTAATA GCTTCTATTA TGATGCCTAA TTGAATGAAT TGGGCAAATG 180  
 GCTCTTTGAT GATAAGTGTG ATAATGAAAA GGGTTAAACT AACATAATC GCATAATATT 240  
 TTTTTCGTTT AATAAGTCGC ACAGGAATGG GCTTCTTTTT AGTTGCTGCA GGAGCATATA 300  
 20 CTGAGATTAC ACCTAAAGAA ATAACTGTTA AAATAATCAT AATTAAAAAG TTAATATGAA 360  
 AATTTACTAT TACTAAAGGT AAAAGTATAA ATAGTATAAT ACTTTCTACA TAACACCAAA 420  
 AAGAAGAAGG TGCATGTGCa CCATGTGCAT GtCTTCTTAT TAAATAAAAT GTTAAATTCTG 480  
 25 TAATTAACGT AAACAGAAAA ATGTTTAAAA TATAGGCAAT AGTATACATA ACAATTAATT 540  
 TACCTATATT TTTAGCTAAG ACCTGCATCC CTAATCGTAC TTGCAAAAAT TGAATATGAT 600  
 CTAAGTTATT TCTCTTTTGA AGATACGTGG CAAACTGGTC AATTTTATTA TCAAAATAAT 660  
 TCAATTTTAC ACCACTCTCC TCACTGTCAT TATACGATTT AGTACAATCT TTTATCATTa 720  
 TATTGCCTAA CTGTAGGAAA TAAATACTTA ACTGTTAAAT GTAATTTGTA TTTAATATTT 780  
 35 TAACATAAAA AAATTTACAG TTAAGAATAA AAAACGACTA GTTAAGAAAA ATTGGAAAAAT 840  
 AAATGCTTTT AGCATGTTTT AATATACTA GATCACAGAG ATGTGATGGA AAATAGTTGA 900  
 TGAGFTGTTT AATTTTAAGA ATTTTATCT TAATTAAGGA AGGAGTGATT TCAATGGCAC 960  
 40 AAGATATCAT TTCAACAATC GGTGACTTAG TAAATGGAT TATCGACACA GTGAACAAAT 1020  
 TCACTAAAAA ATAAGATGAA TAATTAATTA CTTTCATTGT AAATTTGTTA TCTTCGTATA 1080  
 GTAATAAAAG TATGAGTTAT TAAGCCATCC CAACTTAATA ACCATGTAAA ATTAGCAAGT 1140  
 45 GAGTAACATT TGCTAGTAGA GTTAGTTTCC TTGGACTCAG TGCTATGTAT TTTTCTTAAT 1200  
 TATCATTACA GATAATTATT TCTAGCATGT AAGCTATCGT AAACAACATC GATTTATCAT 1260  
 50 TATTTGATAA ATAAAATTTT TTTCATAATT AATAACATCC CCAAAAATAG ATTGAAAAAA 1320  
 TAACTGTAAA ACATTCCCTT AATAATAAGT ATGGTCGTGA GCCCCTCCCA AGCTCGCGGC 1380  
 CTTTTTTGTA ATGAAGAAGG GATGAGTTAA TCATCATTAT GAGACCCGCC GTTAAATAT 1440  
 55

	TCATTTGCAA AGGGCGAAAT GGGTTCTTAC TGAGTTATCT ATTATAAAAA AATAAACATA	1560
	GACTTATGAA AAATCTCTCA TAAATCTATG TTTAGTCATG aCATGTGTTA AATATTATTT	1620
5	CGGGCGCTTC TTATTTATAC AAATCTAATT TAATACTTTT AAATACAGGT ATATTTTCgC	1680
	GTIGCTGTTT TACTTCATTT AAGTTTAAAT CTACAGTCAA AATATCTGCG GATTCATTTA	1740
10	ATTCTCCAAC TAAATCTCCA TTTGGGTTTA TAACTATCGA ATGACCAGCA TATTCTGTGT	1800
	TACCATCGAA TCCAGTGCTA TTAGTTCCAA TGACAAACAT ATTATTTTCA ATTGCACGTG	1860
	CCTTTAGTAA TGAATGCCAA TGTTGAAGAC GTGACATAGG CCATTGCGCC ACATAAAATG	1920
15	CAATTTTAGC ACCACTACGA GCAGGATATC TTAATAATTC TGGAAAACGT AAATCATAAC	1980
	AGATAAGTTG GGTCACATAA GTACCGTCAG ACAATTGAAA GGGTTCAGCT ACGTATTTCG	2040
	CAGCGGTTAA AAATTCATGC TCTCTTAACA TAGGAACTAA ATGAACTTTG TCGTATTCaT	2100
20	TAATCAGCTG GCCACTTTTA TTCACACTAA AAGCTGTATT AAATATTTGA TTGTTTCTAA	2160
	TGTTAGAAAC TGACCCAGCT ACGATATCGA CTTTATATTT TTCAGCTAAA TGTTTAATAA	2220
	ATGAAAACT TTGTCCTAGA TTATTATCTG CTTTTTCATT TAAATGCTCT AAATCATAGC	2280
25	CAITATTCCA CATTTCAGGT AAAACGACTA CATCTACTTC AGCATTCTA TTTTTTTCGA	2340
	ACCATTGCGT TATTTGAGTT TCATTTTGTAG AACTATCTCC AAAACAATC GGTAATTGAT	2400
30	AAATTTGGAC TTTCATAACA TCACATCCTT GATAGATCTT ATATATAACT TACTAAAAGT	2460
	TATGTTGAAA CGCAAAAAAC GAGCACAAGA CATAAAATCA AAGTCCTAGG CTCTACAAAG	2520
	TTATATTGAC AGTAGTTGAT GGGGCCCCAA CATAGAGAAA TTGGAACACC AATTTCTACA	2580
35	GACAAATGCAA GTTGGGGTGG GCTCTAACAT AAAGAAATAC TTTTCTTTA GAAATTAGTA	2640
	TTTCTTATAC ATGAGTTTTA CTCATGTATT CCTATTCTTA AGTGCACATT AGCAGCGGCT	2700
	AATGTGTAAG AACTACTACA TAATGAATAA CTAATGATTC TTTATCATTT CTGTCCCAT	2760
40	CCTAACAATA TATTGATTAT TTTTTTATTA CGAAACGATC TTCCACTGGA TTAAATGTTT	2820
	TTTCGCCAGC AGCTTCACGA ATATCACCAA ATGGCATTG AGCAATAAGT TTCCAACTTT	2880
45	TAGGAATATT AAATTCATTT GAAGTCATCT CATCAACAAG TGGATTATAG TGTTGTAATG	2940
	AAGCACCTAT GCCTTTAGTA GCTAATGCAG TCCAAATTGC AAATTGATGC ATGGCATTG	3000
	TTTGAGTTGA CCATATTGCA AAATTATCAT AGTAGTTTGG CATTTGTTCT TGTAACCAC	3060
50	TTACAACATC TTGATCTTCA TAAAACAAAA TTGTACCGTA TGAATGTTTG AAGTTATCAA	3120
	TTTTTTGTTT AGTTGGCTCG AAATCACGAT TCTCTCCCAT GACTTCTTTT AAAATTGCTT	3180
55	TTGTGTTATC CCAAAATTTA TTATTGTTGT CATTTAACAA GAGAACAATT CTAGTTGATT	3240



	CATCGCTAAT TGATATCGAA TCTTTCAAAT TATATATTGA ACGTCTTTCT TCCATTGCAAT	3360
	TGTCAAAAGT CATTGCTTTT TTATCTTTTT TAAATAAGCC CATAATTATT GCTCCTTCTT	3420
5	TAGTAAAGAA TACTTAATAG ACTAAGTATA AAATTTATAC TCGTACTTGT AAAGCAATAT	3480
	TTACGAAAAT TTCAAGAATA TTAATATTCA TTTTCAAATT CCAAATATAA ATGCATTTTC	3540
10	AACGCATATT TATTATACTT AGATTAAATC TTACATGAAA AAGGGAGGTG TCTCGTGAAA	3600
	TGTCATATCA TTGGTTTAAG AAAATGTTAC TTTCAACAAG TATTTTAATT TTAAGTAGTA	3660
	GTAGTTTAGG GCTTGCAACG CACACAGTTG AAGCAAAGGA TAACCTAAAT GGAGAAAAAC	3720
15	CAACTACTAA TTTGAATCAT AATATAACTT CACCATCAGT AAATAGTGAA ATGAATAATA	3780
	ATGAGACTGG GACACCTCAC GAATCAAATC AAACGGGTAA TGAAGGAACA GGTTCGAATA	3840
	GTCGTGATGC TAATCCTGAT TCGAATAATG TGAAGCCAGA CTCAAACAAC CAAAACCCAA	3900
20	GTACAGATTC AAAACCAGAC CCAAATAACC AAAACTCAAG TCCGAATCCT AAACCAGATC	3960
	CAGATAACCC GAAACCAAAA CCGGATCCAA AACCAGACCC AGATAAACCA AAGCCAAATC	4020
25	CGGATCCAAA ACCAGATCCA GATAACCCGA AACCAAATCC AGATCCAAAA CCAGACCCAG	4080
	ATAAACCAAA GCCAAATCCG GATCCAAAAC CAGATCCAGA TAAACCAAAG CCAAATCCGA	4140
	ATCCAAAACC AGACCCTAAT AAGCCAAATC CTAACCCGTC ACCAGATCCC GATCAACCTG	4200
30	GGGATTCCAA TCATTCTGGT GGCTCGAAAA ATGGGGGGAC ATGGAACCCA AATGCTTCAG	4260
	ATGGATCTAA TCAAGGTCAA TGGCAACCAA ATGGGAATCA AGGAAACTCA CAAAATCCTA	4320
	CTGGTAATGA TTTTGTATCC CAACGATTTT TAGCCTTGGC AAATGGGGCT TACAAGTATA	4380
35	ATCCGTATAT TTAAATCAA ATTAATAAGT TGGGCAAAGA TTATGGAGAA GTTACTGATG	4440
	AAGACATTTA TAATATTATT CGAAAACAAa ATTTCAGCGG AAATGCATAT TTAAATGGAT	4500
	TACAACAGCA ATCGAATTAC TTTAGATTCC aATATTTCAA TCCATTGAAA TCAGAAAGGT	4560
40	ACTATCGTAA TTTAGATGAA CAAGTACTCG CATTAATTAC TGGTGAAATT GGATCAATGC	4620
	CAGATTTGAA AAAGCCCGAA GATAAGCCGG ATTCAAAACA ACGCTCATTT GAACCGCATG	4680
45	AAAAAGACGA TTTTACAGTA GTTAAAAAAC AAGAAGATAA TAAGAAAAGT GCGTCAACTG	4740
	CATATAGTAA AAGTTGGCTA GCAATTGTAT GTTCTATGAT GGTGGTATTT TCAATCATGC	4800
	TATTCTTATT TGTAAGCGA AATAAAAAGA AAAATAAAAA CGAATCACAG CGACGATAAT	4860
50	CCGTGTGTGA TTCGTTTTTT TTATTATGGA ATAAAAATGT GATATATAAA ATTCGCTTGT	4920
	TCCGTGGCTT TTTTCAAAGC CTCAGGATTA AGTAATTGGA ATATAACGAC AAATCCGTTT	4980
55	TGTAACATAT GGATAATAAT TGAACAGCA AGCCGTTTTG TCCAAACATA TGCTAATGAA	5040

AATATTAATG AACTTACTGT TGTAGCAATA ATAAATGCCA CGATACGATT ACCTTTAATC 5160  
 GCATTAAATA ATTCTCCAAA GATTACTTTT CTGAATACAT ATTCTTCTAA TAAAGGACCA 5220  
 5 ATAATAGATA CAAAGAAGAT AAATATAGGT ATTTTTCGAG CAATAATAAT TAGCTTTTCT 5280  
 GTATTAGGAC TTACTTGTTG TCCACCATAA ATTTGCGTTA ATACAATGCT CACTACCATT 5340  
 10 TGATAAATCA TTACCAATGC AAATCCAAGC AATGCCCATG GAATGATATA TTTTITAGGT 5400  
 TCTTTAACTT CTAATTCTAA TTTTGTTGGA TTTTAAATTT TTAAATTAAT TAAAATAATC 5460  
 GTCGTGGCGG CGATTAAAA TAGAACAAGT TGTATGTAA TGA CTGCTTT AGTCAGTTCT 5520  
 15 ATGCCACTAT ATTGTACAAA TGGTAATTTT TTTACAATGA GAAGCGGTAA AAATTGAGAC 5580  
 AATATATAAA TAATAACAGT TAGCAATGAT GCCCATAATC tTGTCATAAT TTTCTCCAA 5640  
 ATATTTGTTT ATAATTTATT TTATCGTAA TAACTTGAAG TTACAAAAC TAATTAAAAG 5700  
 20 GTTATGACTT GAAATTTTGA CCAAATTTGA TTATTATAAA TGTATGTTAG CACTCTTTAA 5760  
 TGTTAAGTGC TAACTTTTAG GTTTTTTAAG GAGGAACAAT CATGCTAAAA CCAATTGGAA 5820  
 ATCGTGTGAT TATTGAGAAA AAAGAACAAG AACAAACAAC TAAAAGTGGn ATTGTTTAAC 5880  
 25 TGATAGTGCT AAAGA 5895

## (2) INFORMATION FOR SEQ ID NO: 2:

30 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 6796 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

TTTGAAAAAA CAAGGTACGA TTGGTTTAAT AACATATATG AGAACCGATT CTACACGTAT 60  
 40 TTCaGATACT GCCAAAGTTG AAGCAAAACA GTATATAACT GATAAATACG GTGAATCTTA 120  
 CACTTCTAAA CGTAAAGCAT CAGGGAAACA AGGTGACCaa GATGCCCATG AGGCTATTAG 180  
 45 ACCTTCAAGT ACTATGCGTA CGCCAGATGA TATGAAGTCA TTTTIGACGA AAGACCAATA 240  
 CCGATTATAC AAATTAATTT GGGAACGATT TGTTGCTAGT CAAATGGCTC CAGCAATACT 300  
 TGATACAGTC TCATTAGACA TAACACAAGG TGACATTAAA TTTAGAGCGA ATGGTCAAAAC 360  
 50 AATCAAGTTT AAAGGATTTA TGACACTTTA TGTAGAACT AAAGATGATA GTGATAGCGA 420  
 AAAGGAAAAT AAATGCCTA AATTAGAGCA AGGTGATAAA GTCACAGCAA CTCAAATTGA 480  
 ACCAGCTCAA CACTATACAC AACCACCTCC AAGATATACT GAGGCGAGAT TAGTAAAAAC 540

55

	AAAGCGTAAC TATGTCAAAT TAGAAAGTAA GCGTTTTGTT CCTACTGAGT TGGGAGAAAT	660
	AGTTCATGAA CAAGTGAAAG AATACTTCCC AGAGATTATT GATGTGGAAT TCACAGTGAA	720
5	TATGGAAACG TTACTTGATA AGATTGCAGA AGGCGACATT ACATGGAGGA AAGTAATCGA	780
	CGGTTTCTTT AGTAGCTTTA AACAAAGATGT TGAACGTGCT GAAGAAGAGA TGGAAAAGAT	840
10	TGAAATCAAA GATGAGCCAG CCGGTGAAGA CTGTGAAATT TGTGGTTCTC CTATGGTTAT	900
	AAAAATGGGA CGCTATGGTA AGTTCATGGC TTGCTCAAAC TTCCCGGATT GTCGTAATAC	960
	AAAAGCGATA GTTAAGTCTA TTGGTGTTAA ATGTCCAAAA TGTAATGaTG GTGACGTCGT	1020
15	AGAAAGAAAA TCTAAAAAGA ATCGTGTCTT TTATGGATGT TCGAAATATC CTGAATGCGA	1080
	CTTTATCTCT TGGGATAAGC CGATTGGAAG AGATTGTCCA AAATGTAACC AATATCTTGT	1140
	TGAAAATAAA AAAGGCAAGA CAACACAAGT AATATGTTCA AATTGCGATT ATAAAGAGGC	1200
20	AGCGCAGAAA TAATATTTTT ATTTCTTAGA TACATTTTAA GATTGTTAAA TAGAATCATT	1260
	AGTGAATCTT ATTTTAAAGA TAGTAAAGGA TTAATCTAAA TAAGTGCGGA TAATATAAAC	1320
25	ATAACAACAT AATTAaMAGA CATAAATGAC aATAAAAGGA GTATAGAAAT GACTCAAAC	1380
	GTAAATGTAA TAGGTGCTGG TCTTGCCGGT TCAGAAGCGG CATATCAATT AGCTGAAAGA	1440
	GGAATTAAAG TTAATCTAAT AGAGATGAGA CCTGTAAAC AACACCAGC GCACCATACT	1500
30	GATAAATTG CGGAACCTGT ATGTTCCAAT TCATTACGCG GAAATGCTTT AACTAATGGT	1560
	GTGGGTGTTT TAAAAGAAGA AATGAGAAGA TTGAATTCTA TAATTATTGA AGCGGCTGAT	1620
	AAGGCACGAG TTCCAGCTGG TGGTGCAATTA GCAGTTGATA GACACGATTT TTCAGGTTAT	1680
35	ATTACTGAAA CACTTAAAAA TCATGAAAAT ATCACAGTTA TTAATGAAGA AATTAATGCC	1740
	ATTCCAGATG GATACACAAT TATCGCAACA GGACCACTTA CTACAGAAAC CCTTGCGCAA	1800
	GAAATAGTGG ACATTACTGG TAAAGATCAA CTTTATTTCT ATGATGCGGC TGCTCCAATT	1860
40	ATTGAAAAAG AATCTATTGA TATGGATAAA GTTTACTTAA AGTCCCGTTA TGATAAAGGT	1920
	GAAGCTGCAT ATTTAAACTG TCCTATGACT GAGGATGAAT TTAATCGCTT TTATGATGCA	1980
45	GTATTAGAAG CTGAAGTTGC GCCTGTAAAT TCATTTGAAA AAGAAAAATA TTTCGAGGGT	2040
	TGTATGCCTT TTGAAGTAAT GGCAGAACGC GGACGCAAGA CATTACTATT TGGACCAATG	2100
	AAACCACTAG GATTAGAAGA TCCAAAGACT GGGAAACGTC CTTATGCGGT GGTTCaATTA	2160
50	AGACAAGATG ACGCTGCTGG TACACTCTAC AATATTGTTG GCTTCCAAAC GCATTTAAAA	2220
	TGGGGAGCTC AAAAAGAAGT CATTAAATTA ATTCCAGGCT TAGAAAATGT TGATATTGTT	2280
55	AGATATGGTG TGATGCATAG AAATACCTTC ATTAATTCAC CGGACGTATT AAACGAGAAA	2340

	TATGTAGAAA GCGCAGcTAG CGGCTTAGTT GCAGGTATCA ATCTTGCGCA TAAAATATTA	2460
	GGCAAGGGTG AGGTAGTATT TCCGAGAGAA ACAATGATTG GAAGTATGGC TTACTATATT	2520
5	TCTCATGCTA AAAACAATAA GAATTTCCAA CCTATGAATG CTAACCTCGG GTTATTACCA	2580
	TCTTTAGAAA CTAGAATTAA AGATAAAAAA GAACGCTATG AAGCACAAGC TAATAGAGCT	2640
10	TTGGATTACT TAGAAAATTT CAAAAAACT TTATAAAATA GTTAGAAAGA CTAGATATGC	2700
	TATTCATTCT TAAGTCATCA ACGAGTAAGT AATGACTTTC TAAATGGAAA ATACTTATCC	2760
	TAGTCTTTTT AATTTTGGAA TTGTTACGTA TTTCTGACAA TTTAGAATTC GCATTCAAAA	2820
15	AATATCTAAA TAAATAACAC GCAATAAGTT GATTGATGTA ACATGTAAGA GAATGTTTTA	2880
	AATAAACTTT ATTTAAAAGG CAATGAAATA ATAAATGGCA AGGCTATTAA TAAAGACTTT	2940
	TAGTAATTAA TTTAAAAAG AGGTATTCTA ATTAACAGGT TTTCCGATTA GTTACAATTA	3000
20	TTTAATTCTC AAAAGATTTA GAATTGATTA TCAAATTACT GTAAGCCCTT TGCTGTATAT	3060
	GCTACAATTC TTATTGATGG AGGGTAAATG TATTGAATCA TATCAAGAT GCGTTTTTAA	3120
25	ATACATTGAA AGTTGAACGG AATTTTTTCGG AACACACATT GAAATCATAT CAAGATGACT	3180
	TAATTCAGTT TAATCAATTT TTAGAACAAG AACATTTAGA GTTGAATACT TTTGAATACA	3240
	GAGATGCTAG AAATTATTTG AGCTATTTAT ATTCAAATCA TTTGAAAAGA ACATCTGTTT	3300
30	CTCGTAAAT CTCAACGTTA AGAACTTTCT ATGAATATTG GATGACGCTT GATGAGAACA	3360
	TTATTAATCC ATTTGTTCAA TTAGTACATC CGAAAAAGA AAAATATCTT CCGCAATTCT	3420
	TTTACGAAGA AGAAATGGAA GCGTTATTCA AAAGTGTAGA AGAGGACACT TCAAAAAATT	3480
35	TACGGGATCG AGTTATTCTT GAATTGTTGT ATGCTACAGG CATCCGTGTT TCGGAATTAG	3540
	TAAATATTAA AAAACAAGAT ATAGATTTTT ACGCGAATGG TGTTACCGTA TTAGGAAAAG	3600
40	GGAGCAAAGA GCGCTTTGTA CCGTTTGGTG CTTATTGTAG ACAAAGCATC GAAAATTATT	3660
	TAGAACATTT CAAACCAATT CAGTCATGCA ATCATGATTT TCTTATTGTA AATATGAAGG	3720
	GTGAAGCAAT CACTGAACGC GGTGTACGAT ATGTTTTAAA TGATATTGTT AAACGAACAG	3780
45	CAGGCGTAAG TGaGATTcAT CCCCACAAGC TCAGACATAC ATTTGCAACG CATTTATTGA	3840
	ATCAAGGTGC AGACCTAAGA ACAGTACAAT CGTTATTAGG TCATGTTAAT TTGTCAACAA	3900
	CTGGTAAATA TACACACGTA TCTAACCAAC AATTAAGAAA AGTGTATCTA AATGCACATC	3960
50	CTCGAGCGAA AAAGGAGAAT GAAACATGAG TAATACAACA TTACATGCAA CAACAATTTA	4020
	TGCTGTAAGA CATAATGGGA AAGCAGCTAT GGCTGGAGAT GGGCAAGTAA CGCTTGGTCA	4080
55	ACAAGTCATC ATGAAACAAA CGGCAAGAAA AGTGCGACGT TTATATGAAG GTAAAGTGTT	4140

	ATTACAACAG TTTAGTGGTA ACTTAGAAAG AGCTGCTGTT GAATTGGCAC AAGAATGGCG	4260
	AGGCGATAAA CAATTACGTC AATTAGAAGC TATGCTAATT GTAATGGATA AAGATGCTAT	4320
5	TTTAGTTGTC AGTGGAAGTC GCGAAGTTAT TGCTCCAGAT GATGACCTTA TCGCTATTGG	4380
	ATCAGGAGGC AACTACGCAT TAAGCGCAGG ACGTGCATTG AAACGCCATG CATCGCATTT	4440
10	GTCTGCTGAA GAAATGGCAT ATGAGAGCTT GAAAGTAGCG GCTGATATTT GTGTCTTTAC	4500
	CAACGATAAT ATTGTTGTCG AAACACTATA ATAATCAGAG CACGATAAAT AATTACGAGC	4560
	AATTAATTTT AGTTAAAAGA CGGAGGAATG AAATTAATGG ATACAGCTGG AATAAGATTA	4620
15	ACTCCAAAAG AAATCGTATC TAAATTAAAT GAATACATCG TTGGACAAAA TGATGCTAAA	4680
	CGTAAAGTGG CAATTGCCCT ACGTAATCGA TACAGAAGAA GTTTATTAGA TGAGGAATCA	4740
	AAGCAAGAAA TTTCACCTAA AAATATTTTG ATGATTGGAC CAACCGGCGT TGGTAAAACT	4800
20	GAAATTGCAA GAAGAATGGC CAAAGTTGTC GCGCGCCAT TTATAAAAGT AGAAGCTACT	4860
	AAATTTACTG AGSTAGGTTA TGTAGGACGA GATGTTGAAA GTATGGTTAG AGATCTTGTT	4920
25	GATGTTTCAG TAAGATTAGT CAAGGCGCAG AAAAAATCAT TGGTACAAGA TGAAGCAACA	4980
	GCTAAGGCCA ATGAAAAACT TGTTAAGTTA TTAGTTCCAA GTATGAAAAA GAAAGCGTCT	5040
	CAAACGAATA ATCCTTTAGA GTCACTTTTC GGAGGTGCAA TTCCAAATTT CGGACAAAAT	5100
30	AACGAAGATG AAGAAGAACC ACCTACTGAG GAAATTAAAA CAAAACGTTT TGAATTAAG	5160
	AGACAGCTAG AAGAAGGCAA ACTTGAAAAA GAAAAGGTAA GAATTAAAGT CGAACCAAGAT	5220
	CCTGGTGCTT TAGGTATGCT AGGTACAAAT CAAAATCAGC AAATGCAAGA GATGATGAAT	5280
35	CAATTAATGC CTAAAAAGAA AGTTGAGCGA GAAGTTGCTG TTGAGACGGC AAGGAAAATC	5340
	TTAGCTGATA GTTATGCGGA TGAAC TAATT GATCAAGAAA GCGCTAACCA AGAAGCGCTT	5400
40	GAATTAGCAG AACAAATGGG TATCATCTTT ATAGATGAAA TCGACAAAGT TCGACGAAT	5460
	AATCATAATA GTGGTCAAGA TGTCTCAAGA CAAGGTGTTT AAAGAGATAT TTTACCTATA	5520
	CTTGAAGGTA GCGTTATTCA AACCAATAT GGTACTGTGA ATACTGAACA TATGCTGTTT	5580
45	ATAGGTGCTG GAGCTTTCCA TGTATCTAAG CCGAGTGAAT TGATACCAGA ATTGCAAGGT	5640
	CGTTTTCCGA TTAGAGTTGA ACTTGATAGT TTATCGGTAG AAGATTTTGT AAGAATTTTG	5700
	ACAGAACCAA AATTGTCATT AATTAAACAA TATGAAGCAT TGCTTCAAAC AGAAGAAGTT	5760
50	ACTGTAAACT TTACCGATGA AGCAATTACT CGCTTAGCTG AGATTGCTTA TCAAGTAAAT	5820
	CAAGATACAG ACAACATTGG TGCACGTCGA CTTCATACAA TTTTAGAAAA GATGCTAGAA	5880
55	GATTTATCAT TCGAAGCACC AAGTATGCCG AATGCAGTTG TAGATATTAC CCCACAATAT	5940

AAATATACAA AAGGAGAAAA ATTCATGAGC TTATTATCTA AAACGAGAGA GTTAAACACG 6060  
 TTA~~CTT~~CAAAA AACACAAAGG TATTGCGGTT GATTTTAAAG ATGTAGCACA AACGATTAGT 6120  
 5 AGCGTAACTG TAACAAATGT ATTTATTGTA TCGCGTCGAG GTAAAATTTT AGGATCGAGT 6180  
 CTAAATGAAT TATTA~~AAAA~~AAG TCAAAGAATT ATTCAAATGT TGGAAGAAAG ACATATTCCA 6240  
 10 AGTGAATATA CAGAACGATT AATGGAAGTT AAACAAACAG AATCAAATAT TGATATCGAC 6300  
 AATGTATTAA CAGTATTCCC ACCTGAAAAC AGAGAATTAT TCATAGATAG TCGTACAAC~~T~~ 6360  
 ATCTTCCCAA TTTTAGGTGG AGGGGAAAGA TTAGGTACAT TAGTACTTGG TCAAGTACAT 6420  
 15 GATGATTTTA ATG~~a~~AAATGA TTTGGTACTA GGTGAATATG CTGCTACAGT TATTGGTATG 6480  
 GAA~~a~~TCTTAC GTGAGAAGCA TAGTGAAGTA GAA~~a~~nAGAAG CGCGCGATAA AGCTGCTATT 6540  
 ACAATGGCAA TTAATTCATT ATCTTATTCT GAAAAAGAAG CGATTGAACA TATCTTTGAA 6600  
 20 GAACTTGGCG GTACGGAAGG CCTATTAATC GCATCAAAAG TTGCAGATAG AGTTGGTATT 6660  
 ACTAGATCTG TAATTGTAAA TGCACTACGT AAATTAGAAA GTGCTGGTGT AATTGAATCA 6720  
 CGTTCTTTAG GAATGAAAGG TACTTTCATT AAAGTTAAAA AAGAAAAATT CTTAGATGAA 6780  
 25 TTAGAAAAAA GTAAAT 6796

## (2) INFORMATION FOR SEQ ID NO: 3:

30 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2073 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

40 ATCCTAA~~A~~AT ThAAAATTAT CACGCCTTTT GaACAGCTTT GTAACCaTct GGACGATCAT 60  
 KAAATTCCaA TGTAATCCT GGT~~T~~TAAaGT TGATCTTTAA CCTTATTTAA AycACCAATT 120  
 GTACGTATAT TATGTTGTTT AGCAAAATCA CGTTTACAG CTAAAGCATA CGTATTGTTA 180  
 45 TACTTCATTG GTTTTAACAT AGTCATTTGA TATTTCTTTT CAAGACTTTG CTTAGCTTGT 240  
 TCATAAACTT TTTTCTCTTC TTTGACTTC AATGGTTCTT TTGTTAATTC ACCTAA~~A~~ACT 300  
 GTTCCAGTAA ATTCTAAATA CCCATCTATA TCGTCAGATT TTAAAGCATT AAATA~~A~~AAAT 360  
 50 GCTGTTTTGC CCATACCATC TTTC~~A~~CTTCT ACAGTATTTT TGGTCTCTTC TTCTATTAAA 420  
 ATTTTATACA TATTTGTAAT AATCGATGGC TCGGAGCCAA GCTTTCCAGC TAACGTAATT 480  
 TTATCACCTT TTTGTGCAAA CATAGGAATA GCGATAGCCA GTATAATAAT CATCACTATA 540  
 55

TCAAAATATAA TTGCCAATAA GGCTGCTGGA ATTGCACCTA ATAATATCAA CGATGCATTG 660  
 TTACGGTCTA TACCTAATAA AATTAAATCT CCTAGTCCGC CTGCACCAAT TAATGCTGCT 720  
 5 AGTSTTGCTG TACCTATAAT TAATACCATA GCCGTCTTTA CACCAGCCAT TATAACAGGC 780  
 ATTGCTATCG GAAGTTCGAC TTTAGTTAAA CGTCTAAATG GTTTCATACC TATACCTTTA 840  
 GCCGCTTCAA TGAGTGATGG ATCAACTTCT TTAATTCCAG TATACGTATT CCTTAAATTT 900  
 10 GGTAACAACG CATACTACT AAGTGCAATA ATTGCTGGCA CACGACCGAT ACCAAATAAA 960  
 GGAATCATT AACTAATAA TGCCAACGAT GGTATGGTTT GAAGAATTGC CGCAATATTC 1020  
 15 ATTACGATTT CAGATATCGT TTTAGTCTTC GTTAATAAAA TACCTAATGG TACCGCAATA 1080  
 GCAGTTGCAA TCAATAATGC GATAAATGAT ATTTGAATAT GTTCTATCAT TGTCGAAAAG 1140  
 AGTTGCCCCT TACGTTCACT CAATATGTCg AAAAAGTTAG TCATGTTGAG CTACCTCCTT 1200  
 20 TTTCTGGGAC AAATATTTGA AGATATCTTT CCTATCAATA ACATATTGAC CTACGCTATC 1260  
 TTCTTGCAATG ACAATGACAC GCTCGCTCTC TGATAAAAGT TGATACAATA CTTCAATTGG 1320  
 TTGATTGTCA TAAACAATTG GATAAGCGCT CATAGATGTA ACCTCATCGA TTGGTTTCAT 1380  
 25 AATATCCAAG TCACGGATAA TTGCGTTCTC TTCAACACAT GGCGCATCAT CTTCTAAATG 1440  
 ACTACCCATA AATTGTTTAA CAAATTCCTT TTGAGGATTA TTTTAAATC CTTCTGGTGT 1500  
 GTCAATTTGT TCAATATGCC CTTCAATCAA AAGACAAATC TTATCACCAA GTTTCATCGC 1560  
 30 CTCTTGAATA TCATGTGTAA CAAATATGAT TGTCTTCTTA ATTTTAGTTT GTAATTCAAT 1620  
 TAAATCATCT TGAAGTTTTT CTCGGCTGAT TGGGTCTAAT GCACTAAACG GTTCATCCAT 1680  
 35 TAAATAACT GGTGGATCAG CTGCTAACGC ACGTATAACT CCTACACGTT GTCGTTGCCC 1740  
 CCCTGACAAT TCATCAGGTT TTCTGTTTTT ATATTTTTCa GGTTCTAATC CAACCATTTT 1800  
 AAGTAATTCA TCTACTCTTT TATCTATATC TTTTCTTTT CACTTTTTCa TTTGTGGCAC 1860  
 40 TTGTGCAATa TTTTCTTTGa WTGTCaTATG TGGGAATAAT GCAATCTGCT GcAATACGTA 1920  
 TCCAATATCC CAACKCATTT CGTATACTGG ATAATCATT ATTGGTTTAT CTTTAAATA 1980  
 45 AATATAACCT TCACTTAAGT GAATGAGTCG ATTAATCATT TTTAATGTCG TAGTTTTTCC 2040  
 ACAACCTGAA GGTCCAATTA GCACAAAAAA TTC 2073

(2) INFORMATION FOR SEQ ID NO: 4:

- 50 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 13321 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- 55

## EP 0 786 519 A2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

	ACTATTCTAG CTTTCATCAGT TATCATATAT TCTTTGAAAC ACTTGTAAGA AAATATAATG	60
5	AGTATTTACT ACATAATGAT ATTTCAAATT AGAAAAAAGG AAGTTATGAT TTAATGGCCT	120
	TGAGCCTATC ATAACCTCCT TTTATCATTT TATTGTTGTG TTGATGTTTC GATAACGTGG	180
	TACATCTTAT CAAACATCAA TTCGAAACCA TGCACCATGG CATCATGATA TTCTTTTTTC	240
10	TTTTGCTTGT ATTCTAAATT AGTAAATCGT CTTTCTTTTT CAACTAATGA ACGATAATAA	300
	AATAGCATTT GGGTGCCACC TGTTCACGT TCAAAAAATT CTACCTCAAT GACATCTTGC	360
15	GTTTCACTTA GTCCAGGCAT ACCGATAGTC ATCTTAACGT ATTCATCCAT AACTAAAGAT	420
	TCATAAATGC CTTCAATCAC ATTTACTTTG CCATTACGTT GTTGATCTAC AATACGATAT	480
	TTACCGCCTT CTTTAACGTC CGCTTCAATC TCTTTATTCTG TTCTGGCTGA TGTCTAAAC	540
20	CATTGTTTCA ACAAATCTTT CTTTGTCCAA GCTTCGTATA CTAACCTCTGG AGAAAATTTA	600
	TAAAGCTTTT CAATTTCAAC TTCGACATGT TCATTCTCTA CATTAAATTT TGCCACTGTT	660
	GTCCACCCAC TTTCGCTCTT ACTTTTATTT TAACGTATTT TTGCTCAGTT CCAAACATAG	720
25	ATGATCATCA TTTTAAAG ATTAGCGTTA TACGGTGAGT ACAACATGAT CTGTTAATAT	780
	AACAAGCCAC CTTACTTGGC TACATCGATA TATTGTTAAG CATTAATGTT TCATTTCTTG	840
30	ACTAGTGTTT TTTTITAGCT TTGAAAATT AAATAAAATC GCAATAAGTC CGCATACACC	900
	TAATAATATA GGATAAATGC TGTATGGGAA TAACATTAAC GGTGAAATAC CAGCTACACC	960
	AGCCGCTGaa ATGACTTGCG GGCTATATGG TAATAAACCT TGGAAGCAGC CTCCAAATAT	1020
35	ATCAAGAATA CTTGCTGAIT TCCTGAATC TACATCATAT TCATCTGCAA TATTTTTAGC	1080
	TAAAGGACCT GACATAATAA TAGAGATGGT GTTGTGTTGCC GTGGCAATAT CTGCGACACT	1140
	TACCAAACTA GCAATTCCTA ATTCTGCGCC ACGCTTTGAT TTCACTTTAG AGCGAACAAA	1200
40	TTGCAACAAC CATTCAATAC CACCATTGTG TTGAATAATA CCGACTAAAC CACCAATTAG	1260
	CAACGCAATC ATAGCAATAT CTTCCATGCT TATAATACCT TTGGACACTG CATCTAGTAG	1320
45	CCCCATCCAA CCGAATGAAC CATCTATGAG ACCAATGATT CCGGCTAATA ATGTTCCGCC	1380
	AATCAATACG ATAATGACAT TTACACCTAA TAATGCTAAT ACCAATACTA AGATATACGG	1440
	TACAACTTTA ATTAGATTAT AATCATAGTt TTTAGCATGA TTTAAAGAAA TGCCATTCTG	1500
50	TAAGAAATAC AGAATAATAA TCGTTAAAT AGCACCTGGC AATACAATTT TAAAGTTTAC	1560
	TCTGAATTTA TCTTTCATTT TCGTATGTTG TGTCTAACC GCAGCAATTG TTGTATCTGA	1620
55	AATCATTGAT AGATTATCGC CGAACATTGC ACCTCCAACA ACTGTAGCCa tTGctAGCGC	1680



	TCCTACAGAC	GTCCCATAG	ATATAGAAAC	AAACATACAA	ATCACAAACA	ATCCTACAAT	1800
	AATTAAATTT	TCTGGGATTA	ATGATAGTCC	TAAATTAAC	GTCGACTTTA	CGCCACCCAT	1860
5	TTTTTCAGCT	GTATTTGAAA	ATGCACCTGC	TAAAATAAAA	ATCAACATCA	TTAAAACAAT	1920
	GTTTGAATGG	CCTGCACCTT	TCGTGAAGAC	CTCAACTTTT	TTAGCAAATG	ATTCTTTTCG	1980
10	ATTCATTAAT	AACGCCACAA	TTACCGTTAT	CGTAATTGCA	ACATTTAATG	GCATTGAAGT	2040
	AAAATCACCT	GTGATAATAC	CTACGCCTAA	AAACAACGCC	ACAAATAATA	ACAAGGGGAA	2100
	TAATGCCCAA	GCATTGCTCT	TTTTATGTAC	TTCCATCCTT	TTTACCTGCT	TTCCAATTAA	2160
15	AAATACCTCT	TTCTCACAAA	CGATGAAGAA	AGAGGTTTTT	ATGTGCTTTA	CCTGCTTATC	2220
	TTCAAACCAT	TACGGTTACT	GGAATTGGCA	CATTTCGAGAT	GTTGCCGAGG	CTTCATAGGG	2280
	CCAGTCCCTC	CACCTCTCTA	GATAAGTGAT	GCTTATTTAC	GTTTACGTTA	CAAGATAATC	2340
20	CTTAGTACGT	CAATCATAAA	TTAATCAGGA	GTCGTATAAT	ATTTTTTCATA	AACAATCATT	2400
	GCTACTGTAA	TAATAATCAA	AACAATAATG	CTAATAACAA	GTAAAAGCCA	CCATTTAAGC	2460
25	ATTAATGCAA	TAAAAATGAA	CACGATAGAC	ACACTTACTA	ATATTAATGA	TATGACTTTA	2520
	AATTGCTGAA	CACGTTGCTT	GGAGATGACT	TTCAACTGTT	TGTTTGATAG	ACGCGTATTT	2580
	TTTATACTGA	TTCCCAGTAT	ATTTTCTAAT	ATTTGAACCA	ATACGATACT	TATTGCAAAT	2640
30	ATAATAATTG	GTAAAACATC	ATAGCTCCCT	ATAGTTAATG	TATAAATTAC	AAATCCAATG	2700
	TAAAGTAACC	CTGAGACAAA	GGATAAAAAG	TATGCGACGT	ATTTGTTAAA	CTTAATGATA	2760
	TGCTTTTTTA	CGTTTTGATG	TGTAAACCAT	ACATTGCAAA	CGATCGCAAC	TGCTACAAAT	2820
35	AATGTGAATA	CTATATATAA	TGGTAATTTT	TGTTCAAGAA	AAACAGTCGC	TATTCCAAAA	2880
	GCTAATGCTA	AAATCAAAAA	TAATATAGCT	CTAGATACTA	TTAATGCCAT	AATAACAACC	2940
	CCTTTGTTTA	ATATCGAGTT	TGCAAATTTA	CGTTTATCAG	CGTTTCTATG	ATCAGTACTT	3000
40	CTACGGGTAG	CGTTTCTATG	TAATTTACAT	CATCTTAACA	TATAAATACT	TCGCTATTTA	3060
	ATTGAAAACA	TATCCTATTA	TTCTTTGTCC	GTTCTGACGT	TTAATATCTA	GCCTTAGGCA	3120
45	TTTCACTTGT	TAATGAATTT	AACTTTCTTC	CACTAACCGT	CCCTAAACCC	AATCCCGCAA	3180
	CAGTTTTTAA	CTTTTTCGTT	GTTGTCCTGA	CATCCTCATT	AAGAAAGTTT	ATTCTGCTTA	3240
	AAACTTATAA	TCCACACCCT	GAGCAAACGC	TCCTTATGAC	AGAGTATTAA	AATAAGCCGA	3300
50	TAAAGATACA	CACCTTTACC	GAATTTTAA	AATACACTTC	ACCAATTTCAT	TTTAATTTAA	3360
	TGGATTGAAG	TAACTAAATT	AATATTATGT	TGTTCAATTA	AAAGCTTCAT	ACAAACCTAA	3420
55	TCTATTTGCA	CTCCACCGCT	AACACCGAAC	ACTTGTCGGG	TTGTATAACT	TGATTCTTCT	3480

## EP 0 786 519 A2

	GTTTTTIGAC CAAATGTTGG GATTTTACTT TGAGGTTGTC CACCAGAAAT TTGTAATGGT	3600
	GACCAGAATG GACCAGGCGC TACACAGTTC ACTCTAATTC CTTTTGGTCC TAATTCTTCT	3660
5	GAAAACTTT TAGTTAATGA AATAATTGCT GCTTTTGAAG CGGCATAATC ATGAAGAATA	3720
	GGACTAGGAT TATAACCTTG TACAGATGAT GTCGTTGTAA TTGACGCACC CGGTTTTTAAA	3780
10	TATTCCAATG CTTTTTGAAC TGTCCAAAAT AGCGGATAGA CATTGCTTTC AAATGTTTCT	3840
	GTAAATGCCT CAGTTGTAAA TCCATGAATA TCATCATGAT ACTGTTGATG TCCAGCAACT	3900
	AAAGTAACAT TATCTAAGCC ACCTAATTGT TGATATGCTT GTTCAACAAG GTCATAGTTG	3960
15	AACTGTTTAT CTCTTATATC ACCAGGAATT AACACTGCCT TTTGACCACT TTCTTCAATC	4020
	ACTTGGCGTA CTCTTGTGC ATCTTGTCT TCACTCGGAA GATAGTTAAT CGCTACATCT	4080
	GCACCTTCTT TAGCATACGC AATTGCTGCT GCACGCCCTA TTGCTGAGTC ACCACCTGTG	4140
20	ACTAATATTT TATAGCCTTG TAAGCGTTGA TGACCTTGGT AAGACGTTTC GCCACAATCG	4200
	GGTGCTGGCG TCATTTTACA TTGTAAACCC GGTACCTCTT GTTCTTGTTT TTCATAATCC	4260
	GTGTTTTTAA ATTTTGTCT AGGATCTTGA GCTGCCATTT TTTTACATCT CCTTATTCGC	4320
25	TTAATGGTTA TTATTTACCC AATCTTCCTA GGAACCTAAT CATGATTACA CTAAAAATTA	4380
	CTTTCTTCTT TATAAAAACA AGCTCGAATT ATTCATGCAA TAGTCTCTTT ACAAATTCAA	4440
30	CAAAATACTC AGGTACTTTT TCCAGAATCC TTTTCATCCGG TTTATATTGA GGATGATGTA	4500
	AATCATATTC ACTATGAGAA CCAATTAACG CAAATACACT TGGAAAAATG TGAATAAAC	4560
	CTGAAAAATC TTCTCCAATC GTAAGCGGCT GTTCCATCAT TCCCACCTTA TATCCAACAT	4620
35	GTGGGGCTAC TGCAATTGCT TTATGCGTCA ATGCTCATC ATTTCATCACA GCGCCAGGTA	4680
	AATGCGTATA ATTTAAATTA ATTTTCATAT TATATGCTTG AGCCAATCCG TCCGCAATAT	4740
	CTTGTAATCG TGTTTCTACA AGCTTTCGTA CCACAGGATC AAAACTACGC ACTGTGCCTT	4800
40	GTACATACGC ATGATCAGCA ATGACATTCC AAGTATTACC ACATGATATT TGTCCAATTG	4860
	TTACTACCGC TTCATCAAAC GCAGATAGAT TTCTACTAAC TATGGATTGA ATACTATTAA	4920
45	TCAATTGCGC CAACACAATA ACTGGATCGT TGCATTGTTT TGGCTTTGCA GCATGACCAC	4980
	CCACGCCTTT AATATGAAAC TCAAAACGAT CTACTGCTGA TGTAAATTGCC CCTGTTTTGA	5040
	TTGCAAATGT ACCTACCGAA CGCGATGGGT CATTATGAAA ACCCAATACT GCTTGTACAT	5100
50	CTTTTAATGC ATGTGTTTCA ATAATTTTAA AAGCGCCATG TCCTAGTTCT TCTGCTGATT	5160
	GAAAAATGAA TTTAACACGC CCAGTAAGAG TGCCCTCAAT TTCTTTTAAT TTTACAGCTG	5220
55	TAGCCAAAAT ACTAGCCATG TGAATATCAT GACCACACGC ATGCATAACA CCTTCATTTT	5280

	CAGCTATACA ACTCAGACCT TGTCCCACTT CAGCAACAAG CCCAGTCGCA AGTGGTAAGT	5400
	CTAATATTCT AATATGATGT TCTGTTAAAA TATCTTTAAT TTTTGTGTGA GTCTTAAATT	5460
5	CTTTATCGGA TAGTTCTGGA AATTGATGAA AATACCTTCT CCAGGTAACA GCTTGATCTT	5520
	TTAATCCCAT CGGTCATTCC CCTTCCTTAA GTCAATGATA TGTTGTCTAC CCTACGATGA	5580
10	TCATCTTTGA CTATTAAACG ATGATTTTAC AACAAATGTAC TCTTGTTAAT TGCTTTCGTT	5640
	AATGATAGAC AGTTGTTTAA TAATATCGTA ACACTGTTGT CAAACTATTG TAACCTTTAT	5700
	AATTGAGACT CTATACAAAA ACGTGTTCTC GAATATACTT GTTTTTACAA ACCACAAAAA	5760
15	GCTCTAAACA TTAGTTTAAA CCAATGCTTA GAGCTTTCTA ATTATTTTAT GCTTTAAAAG	5820
	ATACTGTGTT ATCTACGATG ACCTTACCGT CTTTAATAAC TTTTCTGCG TGATTGATAC	5880
	CAAAATGATA TGGAATATAT TCATGATTTG GTGCATCCCA AATTACTAAA TTAGCCTTAT	5940
20	CACCTGTGTT AATTGTACCC GCGTTAATGT CTATTGCTTT AGCAGCATTG ACCGTAACAG	6000
	CATTCCAAAC TTCATTAGGT GATAGCTTTA ATTTCAAGGC TGCAATCGCC ATAACAAGTT	6060
25	GTAAGTTGTT TGTGACACTA CTACCAGGGT TATAATCAGT TGCTAATGCA ATCGCACCGT	6120
	TATTGTCAAG CATGCCTCTT GCATCTGCAT AATCTTCTTT ACCTAAATAG AACGTCGTTG	6180
	CAGGTAAGAG GACAGCTACA GTATCACTAT TTCGCAACTT TTCTTTTCCT TTATCACTAG	6240
30	AAGCTACTAA GTGGTCTGCT GATATTGCTT GTTCATCAAT TGCTAATTCC AGTCCGCCTA	6300
	ACGGATCAAT TTCATCCGCA TGTATTTTCA CTTTAAAACC TGCTTCTTTG GCTTTTGTGA	6360
	TATAATGTTG CGATTGTTCT ATTGTAAATA CACCTGTTTC ACAGAAAATA TCCGCAAAGT	6420
35	CTGCATATTG TTTTACTTCC GGAAGTAACG CAATCATTTT TTCTAAAAAT GCCTCATTTG	6480
	AACCTGCCTC TTTAGGTACA GCATGAGGCC CTAGGAAAGT ATGTTTCATG TCTAAATCAT	6540
	ATTTCTCAGC TAAACGATTA GACACTTTCA ATTGCTTCAG TTCATTTTCT CTATCTAATC	6600
40	CATAACCACT CTTACTTTCA ACTGCAAGCA CGCCGTGTTT AATCATAGTA AGCAAATCAT	6660
	GCTCTGCTTT TTTAAACAAG TCATCTTCGG ATGTTTCTCT AGTAGCATTG ACGGTAGATA	6720
45	ATATGCCACC ACCCATTTCT AATATTTCAA GGTAAGACTT ACCTTGACGT TTTAATGACA	6780
	TCTCATGTTT TCGAGATCCA CCAAATGTGA AATGGGTATG TGCATCTACT AATGCTGGGG	6840
	ACACTACCTT CCCACTAGCA TCAATCGTCT CAGTCGCATC GTAGTCATCT GTATGTGTTT	6900
50	CAGCATATAC AATTTTGCCA TCTTTAATGA CAACTGTACC ATTTTTCACA ACATTTAATT	6960
	CATCTAATTC CTTACCCTTC AAAGGTTTAT CTGTTGATCT CGGTAAAATT AATTCTGCTA	7020
55	TATGATTAAT TATTAAATCA TTCATTACTT ATCACCTGCT TTATCAATCA TTGGAATATG	7080

	AACACCCATA CCTGGGTCAG TCGTCAATAC ACGTTCCAAT CTTCTTTTCAG CACGCTCTGA	7200
	TCCATCTGCT ACAACAACCA TACCCGCATG AAGTGAATAT CCCATGCCAA CACCGCCACC	7260
5	GTGATGGAAT GAAATCCATG AACCACCTGC AGCTGTGTTA ATGAGTGCAT TCAATACAGC	7320
	CCAATCACCA ACCGCGTCAC TACCATCTTT CATACTTTCT GTTTCACGGT TAGGACTAGC	7380
10	AACTGAACCA GCATCTAAAT GGTCTCGTCC AATAACAATT GGTGCTGAAA TTTCACCGTC	7440
	ACGTACAAGA CGATTTAAAG CTAAGCCCAT TTTCGCTCTT TCTCCATAGC CTAACCAAGC	7500
	AATACGTGAT GGTAGTCCTT GATATGAAAT TTTTCTTCA GCTAAATCAA GCCATCTTAA	7560
15	TAACTTTTCA TTTTCTGGGA AAAGTTTGCG CATTTCTTCA TCCGCACGCT CGATATCTTT	7620
	TGGATCACCA CTCAACGCAG CAAAGCGGAA TGGCCCTTTA CCTTCACAGA ATAATGGTCT	7680
	AATGTAAGCT GGTACAAAGC CTGGGAAGTC AAAAGCATTT TCACTCCGT TATTGAAGGC	7740
20	TACTTGACGA ATATTGTTAC CATAATCAAA TGCTACAGCG CCACGTTTTT GGAATTCAAG	7800
	CATTAATTCA ACATGCTTTG CCATTGAAGC TTGTGACAGT TCAACATATT TTTCGGATC	7860
25	TTTTTCACGC AATACTTTG CTCTTCTAC AGAGTATCCT TGTGGCACAT ATCCATTTAG	7920
	CGGATCATGT GCACTTGTTT GGTCAGTAAT AATGTCAATT TTAAATCCTT TTTCTAGAAT	7980
	CGCTTGATGG ATGTCTACAG CATTTCCAAC TAACCCGATT GATAATCCTT CTCCACGTTC	8040
30	TTTCGCTCT TCTGCTAATT TTAATGCTTC ATCTAAATCA GCTGTTTTAA CATCACAGTA	8100
	TTTCGTATCA ATTGCTTAT CAACACGTGT TTCATCAACA TCCACGCAA TTGCTACCCC	8160
	ATGATTGATA GTAATTGCTA ACGGTTGCGC ACCACCCATA CCACCTAAAC CTGCTGTCAG	8220
35	TGTAACAGTG CCTGCTAAAT CTCCATTAAA GTGTTGATTA CCTAGCTCGG CAAATGTCTC	8280
	ATAAGTACCT TGCACAATAC CTTGAGAACC AATATATATC CAACTACCGG CTGTCTCTG	8340
40	TCCATACATG ATTAAACCTT TTTATCTAA TTCATTAAAA TGATCCCAGT TTGCCCATTC	8400
	AGGCACTAAT ACTGAATTTG AAATTAATAC ACGTGGCGCT TCTTCATGTG TTTTAAATAC	8460
	AGCAACTGGC TTTCCTGATT GTACTAACAT TGTCTCATCT GATTCTAATT CTCGTAACGT	8520
45	TTTCTCTATT GCTTCAAAAG CTTCCTCAATT ACGTGCTGCT TTTCCAATAC CACCATAAAC	8580
	AACTAAATCT TCTGGTCTTT CAGCAACTTC TGGGTCTAAA TTGTTGTATA ACATTCTAAG	8640
	TACTGCTTCT TGTTCCTAAC CTTTACACTC AATACTCAAA CCTTTTTTTG CTTGAATTTT	8700
50	TCTCATAAAA TTGCTCCTG TTCTTTTAAAG AAGTTAATTC CACTAAATTT AAAACGCTTA	8760
	CATTATTATC TTCAATATTC ATTATAGTAT GTTAAAAATAT AGCCAACAAA TATAAATAAA	8820
55	CTAATTATCC ATAGCTTGAA TCTATAAATA AAAGGAGCAA AACACATGAA AATTATTCAG	8880

	CATATTAGCC AGCCATCTTT AACTGCTACG ATTAAAAAAA TGGAAGCAGA TTTAGGTTAT	9000
	GACTTATTTA CACGTTCAAC AAAAGACATC AAGATTACCG AAAAAGGAAT ACAGTTTTAT	9060
5	CGTTATGCGA GCGAATTAGT TCAACAATAT CGATCCACGA TGGAAAAAAT GTATGATTTA	9120
	AGCGTTACAT CAGAACCAAG GATAAAAATT GGGACTCTTG AATCTACGAA TCAATGGATT	9180
10	GCGAATTTAA TTCGAAAGCA CCATTCCGAC TACCCTGAAC AGCAATATCG TTTATATGAA	9240
	ATACATGATA AACATCAATC TATAGAGCAA TTAATATTCA TTTAGCTATA	9300
	ACAAATGAAA AAATAACCCA CGAAGATATA AGATCCATTG CTTTATATGA GGAATCTTAC	9360
15	ATTTTATTAG CACCCAAGGA AACATTTAAA AATCAAAATT GGGTAGATGT TGAAAATTTG	9420
	CCACTCATAT TACCAAACAA AAATTCTCAA GTGCGCAAAC ACTTAGATGA CTATTTTAAT	9480
	AGAAGAAATA TTCGTCCAAA TGTCGTTGTA GAAACAGATC GATTCGAATC AGCAGTTGGA	9540
20	TTTGTTTCATC TCGGCTTAGG TTACGCTATC ATTCCGAGAT TTTATTACCA ATCATTTCAC	9600
	ACGTCTAATT TAGAATATAA AAAAATTCGT CCAAACCTAG GCCGAAAAAT TTATATCAAT	9660
	TACCATAAAA AACGCAAACA CTCCGAACAA GTACATACAT TCGTACAACA ATGCCAAGAT	9720
25	TATTTATATG GACTTTTAGA GGCTCTTTAA CTTAAGTTAT TAGAGCCTCT TATGCAGTTG	9780
	CTCAGTCAAC TGTATACCTT TTGCCTTTAA CTTAAGTTAT TAGAGCCTCT TATGCAGTTG	9840
30	CTCAGTCAAC TGTATACCTT TTGCCTTTAA CTTAAGTTAT TAGAGCCTCT TATGCAGTTG	9900
	CTCAGTCAAC TGTATACCTT TTTCCTTTAA CTTAAGTTAT TAGAGCCTCT TATGCAGTTG	9960
	CTCAGTCAAC TGTATACCTT TTGCCTTTAA CTTAAGTTAT TAGTGCCTCT TATGTAGTTG	10020
35	CGTAGTCAaC TGTaTACCTT TTGCCTTTAA CTTAAGTTAT TAGAGCCTCT TATGCAGTTG	10080
	CGCAGATCAT CGTATAAAAA TTAATGACGT CATTTCAAAA ATCGATACAA AAATAATTTA	10140
40	TTATAAAAAAT TCTAAGAAAG AAGTGAAGCA GATGTTAAAA TCTATTAATC ATATATGCTT	10200
	TTCAGTCAGA AATTTAAACG ATTCAATACA TTTTATAGA GATATTTTAC TTGGGAAATT	10260
	GCTATTGACT GGTAAAAAAA CTGCTTATTT TGAGCTTGCA GGCCTATGGA TTGCTTTAAA	10320
45	TGAAGAAAAA GATATACCAC GTAATGAAAT TCACTTTTCA TATACACATA TAGCTTTCAC	10380
	TATAGATGAC AGCGAATTTA AATATTGGCA TCAGAGGTTA AAAGATAATA ACGTGAATAT	10440
	TTTAGAAGGA AGAGTTAGAG ATATTAGAGA TAGACAATCA ATTTACTTTA CCGACCCTGA	10500
50	TGGTCATAAG CTAGAATTAC ATACTGGCAC ACTTGAGAAC AGATTAAATT ATTATAAAGA	10560
	GGCTAAACCA CATATGACAT TTTACAAATA AGGTGTCAAT ATAAAAAGGC CTCTTGAAC	10620
55	CCGTTAAAAAT TTTAATTAAT TATTATATAA TAAGAGAACT TTTCAAACAA TACAGTTGTT	10680

	TTACTGCAAT TATTTTTTCAA ATATATCAAC GTTAATATAA CTTCTATTAA GAAATACTCA	10800
	CATTCTGCCC TGCAATGCAA ATCTCGTCAC ATATAAATAT TTTTAATTAT TTTAAAAAAT	10860
5	GATGCACTAA ATTAGCAACG AGCTTAGCAG TTCTATTGTC AGCGTCATAT GTTGGATTCA	10920
	TCTCAGCAAT ACTAACTGAA GACACCTTAT CACTTGGAAT AATACGTTTT GCTAATTCAA	10980
10	GAACAGTATG TGGATACAAA CCTAACACTG CCGGCGCACT TACCCAGGC GCAAACGCAC	11040
	TATCAATGAC ATCCATACAA ATCGTAAACA TAATGACATC ATGTTTCATGT AAAAAACGTT	11100
	CAATCATATC TTTAATTGTT GGTGATACGT GACTCAATAA TTCATCTGCA AAGACATAAT	11160
15	CAATCTTTTT CTCTTTAGCA TAATCAAATA AACTTTGCGT ATTACCACCT TGAGCAATAC	11220
	CAAGCACTAA ATAATCTGTG TTTTCATCTT CTTCTAAAAT TTGTCTAAAG CTCGTTCCAG	11280
	ATGTAGATTG TTGTTTCAGCA CGTGTATCAA AATGCGCATC AATATTTATC ACACCAATAG	11340
20	ATTGTGTTGG ATAGACTTTA CGTGTGCTA AATATTGAGC ATACGCAATA TCATGTCCAC	11400
	CACCTAATAA AAATGTTTGT CTATGATTAG CAATTGACTT CGCTGCAAGC ATAGCAAATT	11460
25	CTTTTTGAGT ATCAATTAAT TCCTCATGAT CATGATAAAC ATTTCCGTAA TCGACTAAAG	11520
	TTcACATTGA TTCAAATCCG GCAAACCTGC AAATGCTTGT TTAATCGCAT CTGGTCCTTC	11580
	TTTTGCACCA ATGCGCCCCCT TGTTTTAAAGC AACACCTTTG TCAACAGCAT AGCCTAATAT	11640
30	ACCGACCCCT GATGGCATACT TACTCTTTTC CAGCTTAGAC AAATCTTCAA ATGTTACTGT	11700
	TTGAAATGT CTAAATTTTT TCGGGTCTGT TTCCTATCT AACCTTCCAG TCCATAAATT	11760
	TGGTTCACCT TGCTTGTTACA CAGCATTTCC CCCTCTTATT TATGTGGCTT ATTAACAATT	11820
35	AAAGTATAAC GTATAGGAAA TTTTGAATTC AATTCATAGT TAAATCCGTA TCTTAAAAAT	11880
	ACTTATCTAC ATTACTTTTA CCCCTATTTT CTATGTAATA ACGAATACTT AGCTGATTTA	11940
	TGTTAATAAAA ATACGTCAAG ACTATTACAT TTTCATTAAT ATTGACATAG ACAATTTATC	12000
40	TCTCGGCTTG TAATATGTAT AATTGTTACT AAAAGATATT TTGCTTGTTA CCTAATGGAG	12060
	GTTACATATA ATGAAGAACA ATAAAATTTT TGGTTTTCAA TGGGCAATGA CGATTTTCGT	12120
45	CTTCTTTGTC ATTACAATGG CGTTATCCAT TATGCTCAGA GATTTCAGT CTATAATTGG	12180
	TGTCAAACAC TTTATATTTG AAGTTACAGA TCTAGCACCA TTAATTGCTG CAATCATTTG	12240
	TATACTCGTT TTCAAATATA AAAAGGTCCA ACTTGCAAGT TTAATAATTCT CAATCAGCCT	12300
50	GAAAGTAATT GAACGTCTAT TGCTAGCTTT AATTTTACCT TTAATTATTC TAATTATTGG	12360
	TATGTACAGC TTTAATACAT TTGCAGATAG CTTTATTTTA TTACAATCAA CAGGCTTATC	12420
55	AGTACCTATT ACACACATTC TGATTGGACA TATTCTGATG GCGTTCGTAG TAGAATTCGG	12480

5  
 10  
 15  
 20  
 25

TGTTGTTGGT	TTGATGTATT	CAGTTTTCTC	AGCAAATACA	ACTTATGGTA	CAGAATTTGC	12600
TGCTTATAAC	TTCTTTTATA	CATTCTCATT	CTCTATGATT	CTTGGTGAAT	TAATTAGAGC	12660
GACTAAAGGA	CGTACAATTT	ATATTGCAAC	GACATTCCAT	GCTTCAATGA	CATTCCGACT	12720
TATTTTCTTG	TTTAGCGAAG	AAATCGGCGA	TCTATTTTCA	ATCAAAGTCA	TCGCCATTTC	12780
AACAGCAATC	GTTGCAGTAG	GATACATTGG	TTTAAGCTTA	ATTATCCGAG	GTATTGCATA	12840
TTTAACAACA	AGACGAAACC	TTGAAGAACT	TGAGCCTAAT	AATTATTTAG	ACCATGTCAA	12900
TGACGATGAA	GAAACTAATC	ATACTGAGGC	TGAAAAATCT	TCTTCAAATA	TTAAAGATGC	12960
TGAAAAAACA	GGTGTAGCTA	CTGCATCAAC	GGTTGGTGTT	GCTAAAAATG	ATACTGAAAA	13020
TACAGTGGCT	GACGAACCAA	GCATTCATGA	AGGTACTGAA	AAAACAGAAC	CTCAACATCA	13080
CATAGGTAAT	CAAACCTGAAT	CTAATCATGA	TGAAGATCAT	GACATCACTT	CGGAGTCAGT	13140
AGAATCAGCm	GaATCAGTTA	AACAAGCACC	ACmAAGTGAC	gATTTaACAA	ACGATTCAAA	13200
TGAAGATGAA	ATAGAGCAAT	CATTAnAAGA	ACCTGCGACT	TATAAAGAAG	ACAGACGTnC	13260
ATCAGTTGTA	ATTGATGCAG	AAAAACATAT	CGAAAAAGCT	GAAGAnCAAT	CTTCAGATAA	13320
A						13321

## (2) INFORMATION FOR SEQ ID NO: 5:

- 30  
 35
- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 8549 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

40  
 45  
 50  
 55

ATGTGTTGTA	AACTTTTATG	TTGAAAAAGC	TACTTATCTC	AATGAAAACA	AGTAGCATTT	60
AATAAATTAA	TTAGTATACA	GCTAGTTTTT	CTAATTGTTT	TTTAACTTGA	ATTAAGTTTG	120
ACCGTATTAG	AGAGGCAGAT	TGATCCATCG	TTTGAATTGC	TTGTCCTTCA	TTTTCGTTCA	180
AGCCATTACA	AACAACCTTCA	AACTGTTGTG	CCATTTGATC	AAGACGCGCA	TGAGCTTGTG	240
TGTTTAAAT	AAACATATCG	TCATAATGTG	ATGGCGAATA	GATAATTCGT	CGTTGTATAC	300
AAACGTATAA	AAACCTTGTC	ATATCAACGG	TTTTGGCATT	TTTAAACCTC	TGTGTTTCC	360
ACGCATGTTT	GCCCTTATTT	AAATAATTTG	CCCTTTTTTC	GCCCCGAAAA	AAAAACACAA	420
AAAAATAACC	CACTCCTAA	ATTAATAGGT	GGTGTGGTTT	TGTTGATTGT	AGGGGTATAA	480
AAATAACCGC	ATTATTAAAG	ATACGGTTAC	TCTGTTATCT	GTAAATATAA	TAGTAGTTTA	540

	AAACAGGACT CCACATAAAA ATCAACTCCT TTATATACCA TAATGATACT ATATTTTCTA	660
	GTTTATTTCA ATTTTTCAGT TTTTAAAAAT GAGTTTCTGT TTTTATTTAT ACGCTTTTCT	720
5	GTTTTCTTTT TAAATTTTAT CTTTTTGTTA TTCCATTCAT TGTA AAAATTC TATTAAATTA	780
	ACATAAAAAT TTTTCATGCCC TATTTTATTT GTTGATGAGA TATCAATGTA AAGACTCAAT	840
10	ATTGTTTTTA AATAGATTTG ATGCAACGAC TGATAAACCG TATTACTATC TGCTATGTTA	900
	TTGGTAAAAAT GCATAGAAAA ATATTCTAAT TTATTCATGC AATATATATG GGTTCATTA	960
	TACTTCTTAA TGAGTGTATT TATACCTTGC AATACGTCAT TACTTTTAAT AACAAATTTCT	1020
15	TTTTCACCTG TCGAAAAAGT CCACTGTTTA TCTCCTATAT TTTCTTTAAT TGTTTTCTTG	1080
	TTGTCAAATT CTAAAAATTAT AGCCCGTAAA CACTCTTCTT TATAATTCTC GTTCTTGAAA	1140
	GTACGAAGCA AAATTTTTAT AAATTCGGTA TTGGTGACTT TTTTATAAGT GTGATATTTT	1200
20	GCAATCTCTT TATCAGTAAA GACTGTTCTT AGTTCGTGAT TATCAAAACT TAAATTCATC	1260
	TTATTCTCTA ATTCATTAAT TTTATCTTGC AAACCAACAT TTTCTAAAAT TTTCTTGTTT	1320
25	ATCTCCCCTA TATCAAAACT CCTTTTCGAA ATTAATTTTG AAAACTCGTC TGCCATTTCA	1380
	ACAGCCTTTT CTTTCCTTTT ATACCTTTTG TTAAATTTAT GAACCACCGT TGCAGCATAA	1440
	TACGATATCC CACCAGATAA AATAGATGAT ATTATCGGTA TGTATATATC ACCTTTTCATA	1500
30	TTTCCACCTC TTTTAACACA ATTAAGTATT ATGATACACA ACTTGCGCAA AAAGATGTAG	1560
	ACAGAACATA ATGGCGAACA AAAACAACCA CCCAGTAACT AGTATGGGTG GCGTAGACTA	1620
	TAACAACTCT ATGTTATCAA GATATATGTA TCGAGTGATG GCAAGGAAGA AGTCTCCTGC	1680
35	GGGACCAACA GTCAGATATA TGGCCTCTGC CGGGCTATAT AGTTCACTCC TACTATATAA	1740
	AAGTAAGTAT AACATAAAAA GCACCCCGTA AACTGTTATA CGGGAATGCT AAAGTCATAT	1800
	ATACTACGGG GAGTAGTATG AAAACTATGC TCTCTATCGT AAGAAAAAAC ACCCAGTGAC	1860
40	ATGCTTGGGT GAACAAGGAT AGATGTAAAT AGTTGATGCA TGTGTACACA TCATAACAAA	1920
	AAACTAGCCC GAAGCTAGCT ATAACATAAA AAAATAGGCA AGTACCGAAG TACCTGCCAG	1980
45	TTACGCACAT TTAAATCTTG AGAGTAATGT TAAAAAGTGT ATAGGAATAT TAACATCCAT	2040
	CCAAATAGTT ATTTAATAAC TGTAAGATTC CCTATAATTA ATGTAGCAAA ATTTTATTC	2100
	TAAGTAAATA CTAAATCGTG CTAAACTTAC CAAAACTACT TATTCTATTA CCTGCCTTGT	2160
50	CTACCTCTCC TGTCGCTATA TAACGACGTT GTCCACTATT AGCAATATAA GTAATCCATC	2220
	TATAGCCATT GATGCAATAT GCGCCGTCAT ATTTAATTGT TGCGTTATTA GGTAATACAC	2280
55	CTGTAATTCT TGAATTAGTT GAATAGCCGT CCCTTACGTT ATTACCTTTA ACATTGGCAA	2340



	CTGGCACTGG TGGATTTTTT TGGTTTTTAG CTGATGTTTT AACATTACCA GCTACCAAAC	2460
	CACCTATAGG CTTACCATGA ATCGCACCGG CTATTAATTT AGAATACAAG TCATAGTTTT	2520
5	TCTTAATCCA ATCCATATCA TTTTATTAG TAATAAAACC TAATTCAGAT AAACGATAGT	2580
	TTATATTTAT TTCTGCTGAT ACATTAACGT TTAGTAAATC ATTACGAGGT GTTACACCTC	2640
10	TTATTTGTCC TAAGTTATTT TTAATAACAT CTTGTATACT TTTATCAATA GSTATCTGCAT	2700
	TGAATTGACT TGAAATAATA ACATGCCAC CACTTGCACT TTCTCCTGCT GCGTCTAAAT	2760
	GAATCTCTAG AACAAATGTCA TACCCATGTG ATTTAACCCA ATATAAGCCA TAATCTTTAT	2820
15	TATTTCTTAC ATTAACACCG TAAGCAGTAT CTTGATACAT ATCTTGTGAT TGAATTGAGC	2880
	CACCATATAA TGCAACTTCG TGACCTGCAT GTCTTAAATA CTTAGCGATA TTTGGTGTTA	2940
	TATATTTACG GATAAAATCA CGTTCATTTG TTCCGTTTCC GACTGCTCCA GGATCGTTAT	3000
20	AACCATGACC GGCTACAAGC ATAATTTTTT TAGGTTTAAT TACTGCTTGC TTTTGGCAG	3060
	TTGCTTGCTT AATAACGCTT TTAGCTTTAT CTCCAACACT TACTTTATCT GGGAAATTTA	3120
	ATCTAATAAA ATACATTGGG TCATCGTAAT AATGAACATG TCTTGTAACG GTTTCGGGAC	3180
25	CCCAACCAGG TTGCGCAACG CCATTGTCC AACCTTTACC ATTCCAATTT TGGCCAAACG	3240
	ATGTGAAAGT GTTTAGATTA GCGCTCTCAA CAATTTCAAC ATGTCCaGct CCGCCACCAT	3300
30	ACTTTGACGG GAAAACGACA ATGTCCAAC TTTGCGGTAA AAAGCTATCA TAGTTTTTAA	3360
	TTATTTGCCC GTATTTTTCA ATCCTTGCTT TATTATCAAA TGAATATTA TAAGCGTATA	3420
	AACCTTGTA CcTTTCGCCT GTTGCTATCA TAAAAACAT ATTTGCGTAA TCGTAACACT	3480
35	GAAATCCATA AAACAAATCA GGATTGAACT GCTTCCCTAA TGAATTATCA AACCATTTTT	3540
	CTGCTTGGTT TTTTGTTATC AACATTGGTC AACACCTACC CTAAATCATT TGTGTCGTTT	3600
	ATATTCGTAG GTGTCAATTAC TTCTTTAATT GCGCTTGCC CTGTTGCTTT TCTATACTTG	3660
40	TTTTCAGCTT TATATTTCTT TAGCTTTTGA TTTGCCATT TACCTTCTTG AGATGTTGGA	3720
	TTATCTTTAT ATGTAGTATA TAAAGCAACA ACTGTTAAGA TAATCGATGA AACACTTTCT	3780
45	TCATCTACTG GTATCGGACT TATACCTTTA TTCGCTAAAA ACTGATTGAC TAATGCTAAG	3840
	ATCAATACGA TGTATCTTGT TATTACTTTT GCATCCATTT GTTTGCTCCT TTTATCCAAA	3900
	ATAAAAAGCC AGTGCCGAAG CACTGACTCT TAACTATTAC TTACACTTAC TAAACCAGAA	3960
50	ACACGACCAA AAGCTATATC CTAAAATTCC CTTAAGCATG GTAATCACCT CCTTTAAATG	4020
	CCAAAAATAG TTTTAAACAA GGCTATAACA AATGTACTTA GAATCGTCCC TATTAATCCT	4080
55	AGAATCCACA TCTTGATGTC TCTAATATTT TTAGCATTTT TCTCTTTATT TTTTTCATCT	4140

	TGCGTTCTCA GACTGTCTTC TATTCTGTCG AATTTTTCAA ACATAGTCTT ATCATTCTTCT	4260
	TCTAATCGCG TTAAACGCCA ATCTTGTTTCG TGTGCTTGG TAAATCCAAA CATTACACCA	4320
5	CCCACTTTAT TCAAATTAAG AAGCCATAAG ATTATAACCT ATGACTCTAG ATTTTCTGGA	4380
	TACTTTTCTC CTGTAATAAT TGCATATTC TCTTTATCTA TAACTTCCAT ATCTACATAC	4440
10	CACGCTATAT CTCTTTTACT ATATTCTTTC AATTGATACC ATGTTTTAAT ATCTTCGAAT	4500
	GTTGGTGAAA TTAATTTAAG CATTTTCAGT CTCTCCTTTA ACCTCTTCTA ATTTTTTATT	4560
	AAGTGTCAAC AGTTGTTTTG CCATTAGTGC ATTTTGTCTA TTAAGTTGCA TCGATAACTT	4620
15	TGTACTTTGA ACAACTTGTT TCTGCATACT AGCAACCATT TTTCGTAAGA TGTCATCAGA	4680
	AGCGACTGTG TTTGTTCTT CACTGTCAAT CTGTTGATGC AAGTCATCTT TTTCTTCTGA	4740
	ATAATCTTCG TTAATAACTA TTTCCCATTT TGAATATTTA AAGGCTTTAG GTCTAAAAAC	4800
20	TTGAGAGAAA TTTTCTGGTA AATTTTCAAT ATCAATACCT TCTTCAAAGC CACCAATGAT	4860
	AGCGTATGAA ATTATCTCAT TACGCTTGTT AACTAATATT TGCATTATTT TCTCACTCCT	4920
25	ATAATTTTGT TAATTGTCCC TCTATTGCG TTCGCACCAG AGCCTCTTTG ACTTCCTAAG	4980
	TCGAAATAGA CATCGTTTGA TATAGTTAAA GATGTACGAC TAGATTTAGT TAATCCAAAC	5040
	TCATAAACAC CTCACCATT TCCATCACCA TCTGGAAGAT TTGAGGGATT CAATGAAATC	5100
30	TTTCCTCCTC CAAAAGGACT GCCAACTCT GTAAAGTCAC CACCTGGAAA AGTCCCATAA	5160
	AAAATTAATA AAATAAATTG GTCTAAACTC TCATTTAAGT ACAATGTAGA GCCCACACCA	5220
	TTTGCTGTTT CATCAAAAAT AACCGAATAC CTTTTATTAA ACTTGTCATC TGCGTATAAT	5280
35	TTAGCGTTAC TTTCGGCCAT ATTAGCTTTT GATTGGGCAC TTTGAACAGT TTCAAAAGGT	5340
	GTATTGTAAT CATTAATAGC TAATTCTGAC CACTCAGACC ATGAACCCGC TTCTTTTCTT	5400
	TTAAACAAATA CTTTATTTGT ACCGTTCCGT CGATAAGTCA TACGCTTGTA ATCTGAAGTT	5460
40	ACTACTAAAT ATTCGACAGT ACCGTTAGTA CTAACACCTC TTGGATAATT TATAGCTTGC	5520
	GAAACATAAA TAAATTGGGT TGAATCACCT ATTCTTTGTT CTGGATTATT AAAATCAAAT	5580
45	CCAGTAATCT GCATTATCTT ACCATCATCT TTAGTAATCT TAGCTTTTGT CCAATTTGAA	5640
	GTAGAACCAC TTGTGACTAA ACCACCACTA TCACTGACT GCTTGAAGGC TTCATGTTTC	5700
	TCATCCATAT ATCGCTTTTG CTCATCGAAT GTTCTTGAAT ATGCTTCCGC TTTATTTTCC	5760
50	AAATCAGATA TATGGCTATT AGCAAGTTGC TTTAATTCAT CTATACTTGA AGATTTTGCT	5820
	ATTTGAATAT CTGATAGACC TTTTCTTTA GCTTTTTCAT TCAGACTCGC ATAATCTTCA	5880
55	CCATTTTTTA TAGCCTCGTC CATTGCTTTC GCACGATCCA TAATAGTTTT TTCTAATTCC	5940

	TCAACGTTAA ATGTGATAGT TCTCTCGACA ACTACCACGT CTGAATTACC TAATTCTGCA	6060
	ACCGAAACTT GAGCTTGATA ACTTCCATCT CGTTTAATTA CATCATTAGG TAATTGAAAT	6120
5	TTTAAATAC CTTTAAATGG ATCTAATATT TCTAGTGGAG CAACTACCAT GACTCCTTTA	6180
	CCTCGAATCG CTATTCGTGC kTTGATATTT tCTTCACTCA ATAATAACGG TTGATTATTT	6240
10	TTAGTGATAT TAAAAAGAAG AACAGAAGAA TCACTCTCTC CTGTTCTAAA AGTTATATCT	6300
	AGATTTGAAA TATTTCCATA ATGCGCTGTG TTTTCTAAAT TTATAGCTAC AGATTTCTCT	6360
	AAATTACTCA TTAACCTATA ATTCTCCCTT CGTGTAAGT CCATGGCCCT GAACTTGTTT	6420
15	TACTATCATA ATTTTTCAAT AGTATCTCAG CAGATGCTGT AACACTATTA CGAACTAGCC	6480
	TATGAACAAA GCCACCTGTG TTTGAAGCTT CTACATATAA GTTCCAACCA GCTACCCCTT	6540
	TACGTTCACT TGGAAAATCT GTAAAACGTT TTGTATCATC CGTAGTTAAA TAAAACGACA	6600
20	TGCCTACTAT GTTAATATCT GACATTTTTG TGATGAATGA AGGTACTCTC TCCCATTTAC	6660
	CACTATTTTT AGGCACATAA TTCCAGTCCG AAATGTCTCC AGTTCTTCCA GAAAGCACCC	6720
25	TTTCAAAGT CATCATATTC CTTGCATAAC TATTACGCGT CAATATCTGA ATTACATCAC	6780
	CGCCAGTTTG TGGTGGCTTA ACTTCCAAGA ACCAACCTGC ATCACGCCAT TCTCTTGGTA	6840
	ATGGGAAATC ATCGATTTGA ACTGTATGAT CAGTGTATAA ATAGTAAAGA CCTGGCTCTG	6900
30	TTAACATCCC AAGATTCTTA AGTTTATCAG GCCTCATTGG TAAAGGTTTA ACTCTACCAC	6960
	CTGTGTCACt CaTGATAAAA GGAACGCCTC TTGAGTGAAG TATTTCTAAA ATACCTCTTT	7020
	GCCCAATCAT GAAAATACGA TGTGTTCTAT TTCCaTCACC ACCGACAGTA ACACCTAGCA	7080
35	TCAAAGCTTT TTTACCACTA TCTTTGTCAT AGTATATTTG CAAACCTTtC TgCTTCCGCA	7140
	AATTCGCCAG GAAATGAATC tAgTGTTCCA CCATAGTCAG CATTAACCTG ATACGCTTCT	7200
	TCTCCTGTTT CTAAATCGAA AGCCGTAAA TAGTTTCTAT TATTTGGATT ACTGTCTCCT	7260
40	GTATACCAAT ACAAGTATTT TTCATCAAAA GTCACACCCT GCATTGGTTG GGTTCGTTT	7320
	GTTAGTCTCA TAGGGATACT GATTTTATGC AAAACTTTAT CAATATTTTT ATCAACATCG	7380
45	TCTAACTTC TTATCTCTAT ATAAntCATT GAGTTTTCAA GTTCCCACTG ACTTCTAGGT	7440
	CTCTCaATTC TGTATAGAAT TTTATTTTCT TTTTCATTTA TGACAGGGGT GATGTAGGGT	7500
	TTTTCTGGGT GTCCTGTAAA TACATCTTGC ATACCATACT TGCCATAGCT AATTTCCACA	7560
50	TTAGGCGTAT ACTTGAAACG AACTAATGTA TTCTCATTAT TACCATTTAA GATAAACTA	7620
	TAAATCCATA ACTCATcATC AATATATCTA TAACCGTTAT GTGTACCATG ACCCCCACCT	7680
55	ACAATCAATG AGCTGTCTAT AAATTGACCA TTAGGTCTTA GACGACTTAG CATATAGCCA	7740

ATTACTGCAT TTGTAAGAGG TGCAAGTTCT GTCACAAATA AAAATTCTTG CTTATCAGGT 7860  
 TCAAAACGAT ACTCGATATC AAGAATTTCT TGTTTGGTCT TATTTAATTC TCTTATAGTT 7920  
 5 TCCTCTTTAT TAATTTGAGT TTTGGTTTCC CAATCGTCTA AATGTTCTTT TAATGTGTCA 7980  
 AAGGTTTCGC CGTTTACATT AACTCGAGCT TGAACAATCT CATTAGCACT GTTATTACGT 8040  
 10 GGTGCCACAA CAAGTGCGTT AATTTGACTT TGTAAGATT TGTTTACTGC TGCTTGCGAT 8100  
 CTACCATTAT AATAAATTG CTCAGCGAAG TGTTGAATTG TTTTAGCTyT CTGATGCAAC 8160  
 TTAAACTCTG TTGTCAAGCC AAGCGCAAAT TGCTCTATTC TTTGTAAGTT TTGTATTTCC 8220  
 15 TTAGCTCTAT AATCTCGACC TGCTAAAGCT CCCAAATCCT TTATTAAATA CAAATTTTCC 8280  
 ATAATGCACC TTCCTTTCTA ATAAAATAGC ACTGTACCAA GTTCCCCT ATCGTCAACT 8340  
 GTTATTTTCC ACAATTTACC GTTTGGGGAT TTCTGTACAA TGCTATTTTG AATAATTgcC 8400  
 20 TGctTCGCCT ATTTTAAAT TATCTAATTT ATTTktATCA TTTACCGAAA TGATACCGTC 8460  
 TTGAGGCAAT CCATCAATAn CACTACTGCC TGCATAAGGT ATCCCATTTA TAGCTTTCCA 8520  
 ATGTGTAGCT GGAAAGTACT GTTTATCGT 8549  
 25

## (2) INFORMATION FOR SEQ ID NO: 6:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3601 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear  
 30

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

AGGCGTGTAG TGAATTACGG nTAGGAACT ATGTATCCGA ATGATTTATT GAGACCAAAA 60  
 AGGCATTAAA GTCCATTGAA ATATCnGGTA GCGmGTTGGT ACgTGGACGT GGGGGCCCTA 120  
 40 GATGTATGAG TCAACCATTA TTCAGAGAGG ACATTTAACG TAATAAATTA TAGAmACGAG 180  
 GGTGAAAATA ATGACAGAAA TTCAAAAACC GTATGATTTA AAAGGCAGAT CATTATTAAA 240  
 45 AGAAAGTGAT TTTACCAAAG CAGAATTCGA AGGACTTATT GATTTTGCAA TTACATTAAA 300  
 AGAGTATAAG AAAACCGGTA TTAAGCATCA CTACTTATCT GGAAAAAATA TTGCACTACT 360  
 ATTCGAAAAG AATTCGACGA GAACGCGTGC TCGTTTTACA GTTGCGTCTA TTGATTTAGG 420  
 50 TGCGCATCCA GAATTTTtag GAAAAAATGA TATTCAATTA GGCAAAAAG AATCTGTAGA 480  
 GGATACTGCG AAAGTATTAG GTAGAATGTT CGATGGTATT GAATTCCGTG GTTTTTCACA 540  
 ACAAGCTGTT GAAGATTTAG CGAAGTTCTC TGGTGTACCG GTGTGGAATG GATTAACAGA 600  
 55

	TCTAGAAGGA ATAAACTTAA CTTACGTTGG AGATGGACGT AATAATATTG CGCATTTCATT	720
	AATGGTAGCA GSTGCTATGT TAGGTGTTAA TGTAAGAATT TGTACACCTA AATCATTAAA	780
5	TCCAAAAGAG GCATATGTTG ATATTGcAAA rGAAAAaGCG AGTCAaTATG GTGGTyCAGT	840
	CATGATTACG GATAATATTG CAGArccAGT TGAAAAaTwCm GATGCTATAT ATmCAGATGT	900
10	TTGGGTATCG ATGGGTGAAG AAAGTGAATT TGAACAcGTA TTAATTTATT AAAAGACTAT	960
	CAAGTGAATC AACAGATGTT TGATTTAACA GGTAAAGATT CAACGATATT CTTACATTGT	1020
	TTACCAGCAT TCCATGATAC AAATACACTT TATGGACAAG AAATTTATGA AAAATATGGA	1080
15	TTAGCTGAAA TGGAAGTTAC AGACCAAATC TTTAGAAGTG AACATTCAAA AGTGTTTGAT	1140
	CAAGCTGAAA ATAGAATGCA TACAATTAAG GCAGTAATGG CAGCAACATT GGGGAGTTAA	1200
	TCACTAAATG GAACGATATG AATATGATGT GTCTGATGAT ATAAGTGTCa TGTACAGACA	1260
20	CCTCATATTG GTATTAAAGG AGAAATGAAT ATGAACGAAT CAGGAGATAA CAAACTCAGT	1320
	AAATCTTCTT TAATTGGACT AGTTATAGGA TCCATGATTG GTGGCGGTGC GTTCAATATA	1380
25	ATGTCTGATA TGGGCGGTAA AGCCGGTGGA TTAGCCATTA TTATTGGTTG GATTATTACA	1440
	GCTATAGGAA TGATTTTCATT AGCGTTCGTA TTTCAAAATT TAACCAATGA ACGGCCGGAG	1500
	CTAGACGGTG GTATTTATAG TTATGmTCAA GCAGGATTTG GCGATTTTGT AGGATTTATC	1560
30	AGTGmTTGGG GATATTGGTT CTCAGCGTTT TTAGGCAATG TTGCCTATGC AACACTATTG	1620
	ATGTCAGCAG TAGGTAACCT TTTCCCGATT TTAAAGGAG GCAACACATT ACCAAGTGTT	1680
	ATTGTCGCTT CGTTACTACT CTGGGGTGTC CATTTCTTGA TTTTAAAAGG CGTTGAAACA	1740
35	GCAGCATTTA TCAATAGTAT TGTTACTGTT GCAAAGTTAA TACCGATTTT ACTTGTAATC	1800
	ATATGCATGA TAATTGCATT CAATTTTGAC ACTTTTAAAA CAGGCTTTTT CAGTATGACG	1860
	TCAGAGGGTG TATTGCCATT TAGTTGGGCG AGCACAATGA GCCaaGTtAA AAGTACGrTG	1920
40	CTAGTGACAG TTTGGGTGTT TATCGGTATC GAAGGTGCAG TAATTTTTTC TAGTAGAGCT	1980
	nAAAATGAGA AAGATGTAGG TAGTGCCACG GTTATAGGAC TTATATCAGT TTTAATTATC	2040
45	TATyTCTTAT TAACTGTATT AGCTCAAGGC GTGATTTTGC AAAATCATAT TTCGCAATTA	2100
	GATTTCGCCAA GTATGGCACA GGTGCTTGCA ACTATTGTAG GTGGTTGGGG ATCTACACTT	2160
	GTAAATATTG GTTTAAATTAT TTCGGTACTA GGTGCATGGT TAGGATGGAC ACTGCTTGCT	2220
50	GGTGAATTAC CTTTCATTGT TGCAAAAGAT GGATTATTTc CAAAATGGTT TGCTAAAGAA	2280
	AATAAAAAATG GAGCACCTGT AAATGCACTG CTTATTACCA ATATATTAGT ACAATTATTT	2340
55	TTAATAAGTA TGCTATTTAC ACAGAGTGCG TATCAATTTG CATTTTCACT AGCATCAAGT	2400

CGACAGCAAG CAACTACTAA ACAATGGACG ATTGGTATCA TAGCCTCAAT TTATGCTATA 2520  
 TGGCTTATAT ATGCAGCAGG TATCAATTAC TTATTATTGA CGATGTTACT TTATATTCCA 2580  
 5 GCTCTTCTTG TTTATACaAT CGkTCmAAAG rATwATCAGa CACGTTTGAT TAAATCAGrC 2640  
 TATATTCTTT TTATGATTAT tATCGTACTT GCAGTTATCG GGTAAATTAA GTTATTGATG 2700  
 10 GGAACGATAA ATGTTTTTTA AAAGGAGCGA CAAAAATATG AAAGAGAAAA TTGTCATTGC 2760  
 ATTAGGCGGT AATGCGATAC AGACAACAGA AGCAACAGCT GAAGCACAAC AACAGCTAT 2820  
 TAGATGTGCG ATGCAAAACC TTAAACCTTT ATTTGATTCA CCAGCGCGTA TTGTCATTTC 2880  
 15 ACATGGTAAT GGTCCACAAA TTGGAAGTTT ATTAATCCAA CAAGCTAAAT CGAACAGTGA 2940  
 CACAACGCCG GCAATGCCAT TGGATACTTG TGGTGCAATG TCACAGGGTA TGATAGGCTA 3000  
 TTGGTTGGAA ACTGAAATCA ATCGCATTTT AACTGAAATG AATAGTGATA GAACTGTAGG 3060  
 20 CACAATCGTT ACACGTGTGG AAGTAGATAA AGATGATCCA CGATTTGATa ACCCAACTAA 3120  
 AccAaTTGGT CCTTTTTATA CGAAAGAAGA AGTTGAAGAA TTACAAAAAG AACAGCCAGA 3180  
 CTCAGTCTTT aAAGAAGATG CAGGACGTGG TTATAGAAAA GTAGTTGcGT CACCACTACC 3240  
 25 TCaATCTATA CTAGAACACC AGTTAATTCG AACTTTAGCA GACCGTAAAA ATATTGTCAT 3300  
 TGCATGCGGT GGTGGCGGTA TTCCAGTTAT AAAAAAAGAA AATACCTATG AAGGTGTTGA 3360  
 30 AGCGGTTATA GATAAAGATT TTGCTAGTGA GAAATTAGCA ACGCTGATTG AAGCAGATAC 3420  
 CTTAATGATT CTTACGAATG TAGAAAATGT ATTTATTAAC TTTAATGAAC CTAATCAACA 3480  
 ACAAATCGAT GATATTGATG TAGCAACACT GAAAAAAtAC GCGGCACAAG GTAAGTTTGT 3540  
 35 GGAAGGATCG tGTTGCCAAA AATAGAAGCT GCGtACgtTT GTTGAAaAGtG GGGaAACCAA 3600  
 A 3601

## (2) INFORMATION FOR SEQ ID NO: 7:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

CGACACTATT AAATGAATTA GAGCACAATC TAACAAATCA AATTCATTTT TCAAAAGATG 60  
 AACGACTCAC ACATATCGCT TTAAAGTTAT TCGAAACAAC CGATCCTGTT TCAACAAAGC 120  
 AACTTGCGCA AGATGTTAAT GTTTCGCGTC GGACAATTGC AGATGATATT AAAATGATTC 180

TTATTGGTGA GGAAGATCAT TATCGTAAAG CGTATGCACA CTTTATACAT CAATATATGA 300  
 AACAAAGCTGC ACCTTTTATA GAGGCGGATA TCTTTAATTC AGAATCAATC GCATTGGTTC 360  
 5 GCCGTGCCAT TATTAAGACA TTAAATAGTG AAAATTATCA TTTAGTTCAG TCGGCTATCG 420  
 ATGGCTTAAT CTATCATATA CTCATTGCCA TTCAGCGTTT AAATGAAAAT TTTTCGTTCC 480  
 ATATACCTAT CAATGAAATT GATAAATGGC GACATACTAA TCAGTATGCn ATTGCTTCAA 540  
 10 AAATGATAGA AAACCTTAGAA CGCAGTGTA TGT 573

## (2) INFORMATION FOR SEQ ID NO: 8:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1221 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

TTGATATTTA TAACGTTATA TTTAATAGT TCACCTGGAT TATTAAATAA ATAGTCCGCC 60  
 25 AAATTTTCTT TTTCTTTATC AATCTGaTg TAATTAACaC TTTCGaCTTC TG TAGGAATT 120  
 CTAATGTCAA CAGAAGCATT GATATAAGCT TGATGTTGCA TGCAATCACA CTCCTAATCC 180  
 30 TTCATmTmAA ACGGAGAAGT AAACCCGTCA CTATTCAAAT TCAATCCTTT TGCCCAATCA 240  
 ACAGGCTTAT TCATGATAGT TTCGATTTCC TTAAGTCCAT TTGAACCTCT AGGTATTTCT 300  
 ACAATTACTT CATCATGGAC ATGGCCAACT ATTTTAAAC CTAATGCTTC AAGCCTTGCT 360  
 35 ATAGAAATCG CAAGTAAATC CCTTGCAGTT GCTTGAACAA TATTCTCGAC TAACTTCCCA 420  
 CCATACGTTT TTAACCTTGA CCATTTACGG TTAAGATCTA ACCCCATAAA TTCAACAAC 480  
 TGACTACCCC AACTATTTTC ACCAACTAAA GCTTTTGGAT AAGCTAAAGC TCTTCCACTA 540  
 40 GGCAGTTCAA TCATTAGAAA ACCTTTTTTTC ATATAAAATC TAAGTCCATG TGTATGATGC 600  
 GTCTTTCCGG ATTTTACAGT ATTAATTGCA GCCTCTTGGC AAGCCTTCCA AAAATTAACT 660  
 ATGTTAGGAT TTGCGTTACG CCAACTATCA ACTAAACCTT GTAACCTGTT TTCTTCAATG 720  
 45 CCCATTTCCA ATGCACCCAT TGCTTTTAAA GCTCCAGCGC CACCTTGATA GCCTAAAGCT 780  
 AATTGGGACA CTTTTCCTTT TTGTCTGAGA GGGTCGCCTT TAGTTATGCT TTCTACCGGT 840  
 50 ACATTAAACA TTTGAGAAGC CGATGCTTCA TATATCTTTC CGTGTGTGTT GAATACATCT 900  
 AAACGCCATT GTTCTTTTGC ATACCATGCT ATGACTCTTG CCTCTATTGC AGAAAAATCA 960  
 CTTACTGCTA GTTCATTACC TTCTTCAGCA GTAAATGTCG TCCTAACTAA TTGACTTAAT 1020

AGATCTCTTG CTATTTCTAA TTCAGTATCT GAAATATAAT GCTTTGTTAA ATTCTGAAGT 1140  
 TGTACACCTC TACCTGCCCA TCTTCCAGTA CCGGCACCGT AAAATTGAAA CAGACCTCTT 1200  
 5 ACCCGTTCAT CACTGCACAT C 1221

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1090 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

TTTTGTTTGG TATGAGGTAG CAATGACGAC GTGTCATTGG TGGAGATTGT AAAAATACAT 60  
 20 AATAAAAAGA AGCGGCAATG TATACCGCTC CTTTTTTATA CTACATACCG ATTTTCAACC 120  
 ATCTCTTTCT ACTTAGTAAT AAGACAATAG TATTAACAT AAATAGAAGA ACGAAGAATG 180  
 ATACTATATT TATAATTTCA GTAGGACACA TAAATGTTGA CTCGTTATTC AATATTTTTT 240  
 25 CTACGGCACG ATACATCGTA TTGCTCGCCT CAAATGGAGC AACGATACCA AATATATTTT 300  
 TATTAATGGC AACTAAGATG ACTGAACCAA TCCAATATAC AATGCTGATA CCTAAGCTGA 360  
 30 TTAAAATGTT AGGTGAAACC ATACTAATCG TTCCAACAAC TAAGATATAT TGTAAGATAA 420  
 CGAGTGAAAA TAAGATTATT AATAGTAAGT AATGTGAGAA ATCCGAATAT ATAATTGAAA 480  
 TAATAGTGAT ACTTAGAATT ATGAACACTA AACATTCAAA AAATAACACT GCTACCTTTT 540  
 35 TATAGAAGAA GGTAAAGATA TTATCGCCAA TCAATTTATA AAACAGGATA TTTTATTTCG 600  
 AATACTCTTT ATTAATAAAA TATGCAATAA CAAATGAAAA TAGTAAGAAC CCTAATTGCG 660  
 TTGCAACAGT ATATGAACTG AAGAAAAACT GGCTATAGCT TAAACTTTTA ACTTTGTCTA 720  
 40 TACCTATTGG TAAAAAATAC CCAAGTAAGA AAAGGAATGT GAATAGCACA ACAAGCGTGT 780  
 AAATAATTTT ATTGGAAATA CTTTTTTTAA ATTCTAATTT CAAAGTGGAC ACCTCAATTA 840  
 45 TAAATTAATG TAATCATTTA TGAATCTTC TTTTGATTGG TACTCTTCTA TTTGAAGGTC 900  
 TTTAAAAATA AAGTATTTAC CCGGCAAAGC ACTTAAATCG GATAAATTaT GTGTAATATT 960  
 GATAATAGTT TTAGTTTGAT GGCTTTGAAT AAAATCATTT AAAAATTCAT AAATTCATT 1020  
 50 AACTGTTTTT TTGTCTAAAG CGTTTGTAAC TTCATCTAAT ATGATTAAAT CATGATCTTC 1080  
 CAATAAGAAA 1090

(2) INFORMATION FOR SEQ ID NO: 10:



(A) LENGTH: 904 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

10	TTAGGACTAT TTTATCATAT TCATTTAAAT TACGGCTAAA AATTTTAAAA ACGGGGATTA	60
	ATATATGGAA TTAAGCTATG AAAGTTAATT GATACTTGCA TTTTACGCTG ATTTATATAA	120
	GAATAACTAT TGTATAGTTT TAAAAACGAA CGTACGTTTG CAGGAGGCGA AATCATTGGC	180
15	AATGAATAAA CAAAATAATT ATTCAGATGA TTCAATACAG GTTTTAGAGG GGTTAGAAGC	240
	AGTTCGTAAA AGACCTGGTA TGTATATTGG ATCAACTGAT AAACGGGGAT TACATCATCT	300
	AGTATATGAA ATTGTCGATA ACTCCGTCGA TGAAGTATTG AATGGTTACG GTAACGAAAT	360
20	AGATGTAACA ATTAATAAAG ATGAGTAGTAT TTCTATAGAA GATAATGGAC GTGGTATGCC	420
	AACAGGTATA CATAAATCAG GTAAACCGAC AGTCGAAGTT ATCTTTACTG TTTTACATGC	480
	AGGAGGTAAA TTTGGACAAG GCGGCTATAA AACTTCAGGT GGTCTTCACG GTGTTGGTGC	540
25	TTCAGTTGTA AATGCATTGA GTGAATGGCT TGAAGTTGAA ATCCATCGAG ATGGTAATAT	600
	ATATCATCAA AGTTTTAAAA ACGGTGGTTC GCCATCTTCT GGTTTAGTGA AAAAAGGTAA	660
30	AACTAAGAAA ACAGGTACCA AAGTAACATT TAAACCTGAT GACACAATTT TTAAAGCATC	720
	TACATCATTT AATTTTGATG TTTTAAGTGA ACGACTACAA GAGTCTGCGT TCTTATTGAA	780
	AAATTTAAAA ATAACGCTTA ATGATTTACG CnwGGgTAAA GAGCGTCAAG AGCATTACCA	840
35	TTATGAAGAA GGGAtCaAG rGTTgTTAGT atGTCCAaTG ArGGAAAAGA AGTTTTGCCT	900
	GACG	904

(2) INFORMATION FOR SEQ ID NO: 11:

40

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 11271 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

50	GATTTCTAAA TCAAGATCTG TTTTACGATA ACCATTCAAA CCTTGACGTT CATCTTCTTC	60
	AGGTTGATTT TGTTGCTGTG TGTCTTTGTT GTCAGAAGTC GCTACTGTTT TTTTATTATC	120
55	TGTTTCTTTA GTCATAACAA ACGCCTCCGT TATAAACGC TATATTTAAT GATATGTGAT	180

	TTAATAAGAC GATTGAGCAA GTTTTAAAGT ATTATTTGAC TATGTTGGAT TAGGCATCTA	300
	GTCCTATAAT ATCACTGACA TTGTCAAAAT GATGATCTTT TAAGTAACGT GCGATGCCTT	360
5	TGTTTCATTTT CTTAGTTAAA CCTGGGCCTT CAATAACAAG TGATGAATAA ATTTGAATAA	420
	GTGACGCACC GTGACGCATC ATTTTGATTG CATCTTCAGT ACTGAATACG CCGCCTGTAC	480
10	CTATAATTAA AAATTCACCA TTTGTTTGCT GATAAGCATA CTTAATCAAT TTTAAATTAC	540
	GTTCAAATAA TGGACGACCA CTCAAACCGC CTTCTTCGAC TTTATTAGCA GAAGTTAAAC	600
	CATCTCGTTG TCGCGTTGTG TTTGCTAAGA TGATACCGTC AAATGTCTCA GTAATCGCTG	660
15	GTAATAGTGC TTTTAAGCCA TCGAAATCCA TATCAGACGT TAGTTTTAAA TAAATTGGCA	720
	CTGTTACATC ATGTTGTTTT TTAATGCTG TTAAGCTTG GCATAACATT GAAAATTCAT	780
	CTTTATCATG GAAGTTTTGA AGATTTTCAG TATTTGGAGA ACTGATGTTG ACTGTGAAAA	840
20	ATGAAACGTC GTGTTTAAAC GTATCAATAA CCTTTATATA ATCTTGATAA CGCGCTTCAT	900
	AAGGTGTCAT TTTATTCACA CCAACATTGA TACCAACAGG TACTTGATAA GCATTTTTAC	960
	GCAAATGACT TAGTGCTTTG TTCATACCAA TATTATTGAA GCCCATTCGA TTTATCAAGG	1020
25	CGTCATCTTC TAATAATCTA AACATGCGTG GTTGAGGGTT ACCCGGTTGA GGTTTAGGTG	1080
	TGATACCACC TAATTCTAAA GCACCGAATC CAAGGTGTTT CAATGCTTTT GGTACTTCGC	1140
30	AAGATTTGTC GAAACCAGCT GCTAAGCCAA TTGGATTGTC GTACGTATTA CCTTGTATCG	1200
	TTTGTGATAA CGTTGGATTG TTATAAGTAA ATAGTTTATC GACGACTGGG AATAAAACCG	1260
	GAAACTTTTG TaACGTTTTT AATGCATCGA TAGTTAGTCC GTGTGCTTTT TCGGGTTCGA	1320
35	TTTTGAATAA GAAAGGTTTA ATTAATTTGT ACATGAGTAT GCTCCTATTT CATTATATTT	1380
	GAGGCTTACT ATCCTCAACT TAATATATGT GAAATATATT CTTTTAATAG ACTAGCATT	1440
	CCATACATAA TTTCTAGTT AAAACTAAAA AGTTTTGAAA ATTGACGCAA gTTTGAATAA	1500
40	CGTTTTTAAG ATTAAATCAT CCTAATTAGG CAATATTATA GTATAAAGTA AGTAGATTGG	1560
	AAGGTGTTTG TATGAATGAA CAATGGTTAG AGCATTTACC TTTAAAAGAT ATTAAGAGA	1620
45	TTTCACCACT GAGTGGTGGT GATGTAAACG AAGCATATCG AGTCGAAACA GATACGGATA	1680
	CATTTTTCTT ACTTGTCCAA CGTGGACGTA AAGAATCATT TTATGCTGCA GAAATTGCAG	1740
	GTTTAAATGA ATTTGAACGT GCAGGTATCA CGGCACCTAG AGTAATTGCA AGTGGCGAGG	1800
50	TTAACGGTGA TCGGTATTTA GTGATGACGT ATTTAGAAGA AGGGGCTTCA GGGAGTCAAC	1860
	GCCAAATTAG GCAACTCGTA GCTCAATTAC ACAGTCAGCA ACAAGAAGAA GGCAAATTTG	1920
55	GCTTCTCATT ACCTTATGAA GGTGGCGATA TTTCTTTTGA TAATCATTGG CAAGACGATT	1980

	GGCTATGGGA TGCCAACGAT ATCAAAGTAT ATGACAAAGT GCGACGTCAA ATTGTGGCGG	2100
	AATTAGAAAA GCATCAAAGT AAACCGTCTT TATTACATGG TGACCTATGG GGTGGTAATT	2160
5	ATATGTTCTT ACAAGATGGT CGTCCGGCGT TATTTGATCC AGCGCCATTA TATGGTGACA	2220
	GAGAATTCGA TATCGGTATT ACAACGGTAT TTGGTGGTTT TACGAGCGAA TTTTATGATG	2280
10	CGTATAATAA ACATTATCCA CTCGCAAAAG GTGCATCCTA TAGACTTGAA TTTTATCGTT	2340
	TATATTTATT GATGGTCCAT TTATTGAAAT TTGGTGAGAT GTACCGTGAT AGTGTGCGC	2400
	ATTCTATGGA TAAGATTTTA CAAGATACAA CAAGTTAGTT AAGACGTTAG ATTGAGATAA	2460
15	ATAGATAATA TGCACAGATA TTTTACAAT GAGAAGCGAT ACAGCTGCCT CAATAAAAAT	2520
	ATTTGTGCGT TTTTATTGTT GGAAAATAAA ATTTTAATCG CTATTGTTAA TTTCTGTAAT	2580
	GTAAAACAAG GTTGAGTTAC AATAAAAGTG ATTTTATAAC TTTTGTTCATAATAAATTCT	2640
20	AGGAATGATA CATATTTATT GATACAATAA TTTTGAATAT AATCATAAAA CAATATTTAA	2700
	GTATAATTGA ATGTTTGAAT ATCATATATT GATACAGTTT CTAATAATTT TAAAATAATT	2760
	TAAATGGAGA GAGGTGTAAA TGATGAGTAC AGTTCAAAGT GATATTTTAA AGACCAATAG	2820
25	TGCATCATCA TCTATTAAAA GCGCTGTTGA AACATGTAAT AATGTGTCGA AACCGGATAA	2880
	AGATGAAAGT ACAACAGTAA GTGGAAATAA TAATGCTCAT AGTGTGATAG ATGATTTGAT	2940
30	GAGTAAGAAT CAATCTGTTG CTGAAGCAAT ACGAACTGCG AGCGATAATA TACAAAAAGT	3000
	TGGTGAGGCT TTTGACCAAA CTGACGTAAT GATTGGTAAT GAAATTGGTA AAAATTAAAA	3060
	CGTGGTGAAA TGATGTCGAA TAAACTGGAT GAAATCAATA AAATAATCAC AGCGAAACAT	3120
35	GAGCAAATGG ATGACTTATA TGATGAAAAG CGAGAGGTTA AAGCATTGAT AGATGAAAGT	3180
	GATGCGCTTA ATCATTCGAT AGATCAATTA TATCAACATT TAGGTGAGCG TTATTATAGT	3240
	AGCAATATGG CTAGTCGTAT GGAACAGTTC CGCGATGAAT TTCATTTTGC GAAACGACGT	3300
40	TCAACGGAAG CGTTATACGA GCAGCAACAG CAAATTCAAC ATGGCATTTCG TAAAGTGGAA	3360
	GAAGAGATGA TTGACTTGGA AATGCGAAGG AATGTTGAAA TTGAGACGGT GACAAAGGAG	3420
45	GAAAATAAAT GGAAACAATA GGAAGCATT TTTATTTAAA AGAAGGTTTCG CAAAAGTTAA	3480
	TGATTATTAA TAGAGGmCCA aTTGTAGAAA TTGAAAATCA AAAGTATATG TTTGACTATT	3540
	CTGCATGTAA ATATCCGATT GGTGTTGTAG AAGATGAAAT TTATTATTTT AACGAGGAAA	3600
50	ATATAGATTC AGTTATTTTT AAAGGTTATT CTGATCAAGA TGAGGTTAGA TTCAAGAGT	3660
	TGTTTTGAAAA TATGAAACAA AATTTGGATA GTGAAATACA ACGTGGAGAA GTTACACAAC	3720
55	AATAAAGAAA TACTTTTTCT TTATTGGGGT GGGACGACGA AATAAATTTT GTAAAAATAT	3780

	ATGTCATTCA TAATCATTTG AACTAAACGT AGCAGCCTTA AATTTTAAAA AAAGACACAT	3900
	ACCAACTTCC GAAATGTAGA TGAATTCCTC ACAATAACGG AAGTTTTTCT TTTAATATTG	3960
5	AAATTTCTCA AGGATAGGTC TATACTTTAT AAATCGTAAT TATTACGATT TATAATCAAA	4020
	AACAATAACT TGAAATAGAT CATTGAGGGA GTGTTAATAT GCAACATCAT AAAGTGGCTA	4080
10	TTATcGGTGC CGGTGCTGCA GGTATAGGTA TGGCCATTAC CTTAAAAGAT TTCGGTATAA	4140
	CAGATGTCAT TATTTTAGAA AAAGGAACAG TAGGACATTC ATTTAAACAT TGGCCGAAAT	4200
	CGACCCGTAC GATCACGCCA TCATTTACGT CTAATGGATT TGGCATGCCT GATATGAATG	4260
15	CAATTTCCAT GGATACTTCA CCAGCATTTA CATTTAATGA AGAACATATT TCCGGAGAAA	4320
	CATATGCTGA ATATTTACAA GTGGTTGCCA ACCATTACGA GCTGAATATC TTTGAAAATA	4380
	CAGTTGTCAC AAATATATCT GTAGATGATG CATATTATAC GATTGCAACG ACAACAGAGA	4440
20	TATATCACGC GGATTATATC TTTGTGCGAA CAGGTGATTA TAATTTCCCT AAAAAgCCAT	4500
	TTAAATATGG TATTCATTAT AGTGAAATTG AAGACTTTGA TAACTTTAAT AAGGGGCaAT	4560
	ATGTGGTTAT CGGAGGTAAT GAAAGTGGCT TTGATGCTGC ATATCAACTT GCAAAAAATG	4620
25	GCTCTGACAT CGCACTTTAT ACTAGCACAA CCGGTTTAAA TGATCCGGAT GCTGATCCTA	4680
	GTGTTAGATT GTCACCTTAT ACACGTCAGC GACTAGGTAA TGTCATTAAG CAAGGTGCTC	4740
30	GCATCGAAAT GAATGTACAT TATACAGTTA AAGATATTGA TTTAACAAT GGACAGTATC	4800
	ATATCAGTTT TGATAGCGGA CAAAGTGTCG TTACACCTCA TGAACCAATA CTAGCAACTG	4860
	GCTTTGATGC AACAAAAAAT CCAATCGTTC AACAAATTATT TGTGACAACA AATCAAGATA	4920
35	TTAAATTAAC AACACATGAT GAATCGACAC GTTATCCGAA TATTTTTATG ATTGGTGCAA	4980
	CAGTTGAAAA TGATAATGCC AAATTATGCT ATATCTATAA ATTTAGAGCG CGATTGCGAG	5040
	TACTTGACACA TCTTTTAACA CAGCGGGAAG GcTTACCAGC TAAACAAGAT GTCATTGAAA	5100
40	ATTATCAAAA AAATCAAATG TATTTAGATG ATTATTCATG TTGTGAAGTG TCATGCACAT	5160
	GTTAGAAGTG AAATATGATA TGAGAACTGG GCATTATACG CCCATACCTA ATGAACCTCA	5220
45	TTATTTGGTT ATTAGTCATG CGGATAAACT TACCGCAACA GAAAAAGCGA AATTAAGATT	5280
	ATTAATCATA AACAGAAAT TAGATATTTT ATTGGCAGAA AGTGTAGTTT CTTcGCCTAT	5340
	AGCGAGTGAA CATGTGATAG AACAAATGAC ACTATTTCAA CATGAGCGAC GACATTTAAG	5400
50	ACCTAAAATA AGTGCGACAT TTTTAGCCTG GTTGTTGATA TTTTAAATGT TTGCATTGCC	5460
	AATCGGTATC GCTTATCAAT TTTCAGATTG GTTTCAAAAT CAGTATGTGT CAGCATGGAT	5520
55	AGAATATTTA ACTCAAACAA CATTGCTCAA TCACGATATA TTACAGCATA TATTATTTGG	5580

	ATTGATTAGT TTATCAACTG CTATAATTGA TCAAACAGGA CTCAAATCAT GGATGATATG	5700
	GGCAATTGAA CCGTCAATGT TATGGATAGG ATTACAAGGT AATGATATCG TGCCACTATT	5760
5	AGAAGGGTTT GGATGTAATG CAGCAGCTAT TTCACAAGCA GCACACCAAT GCCATACCTG	5820
	CACGAAGACA CAGTGTATGA GTTTAATAAG CTTTGGTAGT TCTTGTAGTT ATCAAATAGG	5880
10	TGCGACATTA TCTATTTTAA GTGTAGCTGG AAAGTCATGG CTATTTATGC CGTACTTAAT	5940
	ATTAGTACTT TTAGGTGGCA TCTTACATAA AGGATATGSGT TGAAAAAGAA TGATCAACAA	6000
	CTTAGCGTTC CGCTACCTTA TGATAGGCAA TTACATATGC CAAATATACG TCAAATGTTG	6060
15	CTACAAATGT GGCAAAATAT ACAAATGTTT ATCGTTCAAG CGCTACCTAT TTTTATCACA	6120
	ATCTGTCTTA TTGTTAGTAT TTTATCACTA ACGCCAATTT TGAATGTTTT ATCACAAATA	6180
	TTTACACCTA TATTATCGTT ATTAGGCATC TCGTCAGAAT TGTCACCAGG GATTTTATTT	6240
20	TCAATGATTC GAAAAGACGG CATGCTCTTG TTTAATTTGC ATCAGGGCGC CTTATTACAA	6300
	GGAATGACAG CAACACAGTT ACTACTACTT GTGTTTTTTA GTTCAACATT TACAGCGTGC	6360
	TCGGTCACAA TGACGATGCT TTTGAAACAT TTAGGTGGTC AGTCAGCACT AAAATTAATT	6420
25	GGAAAGCAAA TGGTGACATC ATTGTCTTTA GTTATTGGTG TAGGCATCAT TGTTAAAATA	6480
	GTAATGCTGA TTATTTAAAA AAAATGAACT ATAAGTGAAT ATAGAGTCAT GTCAGTCAAT	6540
30	AGGAGATCTA TCTTGGAATA TGCTATTCTA ATGAAGTATA AGAGGAGAGT CGCAGATGAA	6600
	AATAGTTATT ATAGGTGGGT TTTTAGGTGG CGGTAAAACG ACTGTCTTAA ATCATTGTCT	6660
	CGCTGAATCA TTAAAGGAAT CGCTGAAACC AGCAGTCATC ATGAATGAAT TTGGGAAAAT	6720
35	GAGTGTGAT GGTGCCTTAG TATCTGAAGA CATACTTTA AGTGAAGTGA CAGAGGGGTG	6780
	TATCTGTTGT GCAATGAAAG CAGATGTATC AGAACAGTTA CATCAATTAT ATTTAAAAGA	6840
	GCAAACAGAC ATTGTATTTA TTGAATGTAG TGGGATTGCA GAACCGGTCT CTGTCTTAGA	6900
40	TGCTTGTTTA ACGCCTATTT TAGCTCCGTT TACAACAATT ACACATATGA TTGGTGTAAT	6960
	AGACGCAAGC ATGTATAAAC ACATTAAATC ATTCCTTAAA GACATCCAAG GCTTATTTTA	7020
45	TGAGCAATTA GCATATTGTT CTGTCTTATT TGTTAATAAA ATAGATTCAG CAGATGTTGA	7080
	AACAACGAGC AAATATTGA AAGATTTAGA AGTTATTAAAC CCAGAGGCCG ATATACAAGT	7140
	CGGTATGCAT GGCAGCGTCA CTTTGCCAAT ATCAGTTAGA CAAATGACAG CAACTTCTGA	7200
50	CAATAAACAT AAGTCTTTAC ATCAAATGAT TAATCATCAA TTTGTGCAAT CACCAGTCAA	7260
	ATGTACTAAA GCAGAGTTTA TAAAACGTTT AGCATGCCTT CCGTCTCATA TTTATAGGTT	7320
55	GAAAGGGTTT ATGACATTTG AAGACACCGC ACATACGTAT CTCATTCAAT TTACACAAGG	7380

	CGGAAAGGGT	ATTTCAAAAG	AAGACTATCA	ATGTTTGGAA	CAGTAGTGTT	TTCAGTGGAA	7500
	GAGAATGGTT	AACATGCCTT	CATGTATAAT	AACGAGTTGA	TTTGAACGTT	TAAGCGTAAA	7560
5	TAAAAATAAG	CTTGGTCAGC	CATCAAATAT	AATTTGAAAA	CTGTCCAAGC	TGTTTTATTA	7620
	GAGAACAATC	AATTAACCCC	ACATATTTAA	TAATACATCA	GCAAAGCCTT	CAGGTTTTTG	7680
	AATATAACCT	AAGTGACCGC	CTGGAATATC	TACAATAGGT	ATGCCAGTTT	CTTTATTTAT	7740
10	ATAAAAGTTA	ACATCTTGTT	GGAAGGAGCC	TCTAGAATCT	GTCCCATTTA	GTAGGGTGAT	7800
	TTTATCGCTG	TATTTTGTGA	AATCATCCAA	AGTAATATCT	GAATGCGTAT	ATTGTCTAAT	7860
	TTCAAATTCT	GACCAGAACA	TCGTACGTTT	GTAAGTTTCT	ATACGTCCTT	CTTCAGTATC	7920
15	AGCAGGTTGA	GACATCATTT	TTGCATCAAT	TGGTGCAGTA	TTTAATGTTT	CGCCAAATGT	7980
	TTTCATGCCT	TTTTCTAAGC	CTTCTGTTAA	AATTTGATGC	ACAATGTCAT	CATTTTTATC	8040
20	TTTCCAATAA	GTAAGTCTGT	GTAAAAATGT	ATTAATTGGT	GGTTTCGTGA	ATGCAATCTT	8100
	TTTAACGACT	TCAGGGTAAT	CTTTTAACAC	ATGCATCGCA	ACGATTGAAC	CTGAACCTGA	8160
	ACCTAATATA	TAGACAGGTT	CATCACTTAA	TGACTTTGCA	AGTTCGGCAA	TGTCCTGTGC	8220
25	GTCGCGTTTG	ACACGATAAT	CACTGTCAGG	GTTTGAAGCG	GAATCAGGGA	GTGGTTCAGT	8280
	TAACTCGCTT	TCTCCATAAT	CACGACGATC	AACGGCTACA	ACAGTAAAAAT	GGTCTTTTAA	8340
	CTGTTCTGCA	AGAGGCAGAA	AAATGTCTCC	GGTACCGTTT	GCACCAGGAA	TAAAGATGAG	8400
30	CACGGGTCCT	TGTCCGACTT	GGTGGTATCG	TAATTTAGCG	CCTTGTAATT	CTAAAGTTTC	8460
	CATATTCAAT	GACCTCCATT	TGTTAATTGT	TAGGTGATAA	ACCTAATAAT	TTAGCACCAT	8520
35	TTGTATAACT	TATTTTCTCT	TTTTCTTCAT	CTGTTAAACC	CAGTTCATCT	AAAAATACAC	8580
	CTAATTTTTTC	AGGCTCAATA	TATGGATAAT	CAGCAGCATA	AAGAATTCTA	TCAATACCTA	8640
	CTTCTTTCTT	GACTAAATCA	AACTGTGGCT	TCGTTAACAT	GCCACTCGGT	GTGATATAAA	8700
40	AATTATTTTT	AAAGTAATAG	CTTACAGGGT	GGTTCAAATG	TTCAGCGAAT	AAAGCTTCAT	8760
	CCATACGTTT	TAAGAAGAAT	GGGATAAACT	CACCCCAATG	TCCAATAATC	ATATTTAACT	8820
	TTGGATAACG	ATCAAAAATA	CCAGATAATA	CTAGATGTAT	TGTATGAATG	CCGACATCAA	8880
45	TGTGCCAACC	ATAACCAAAA	CAAGCAAATG	TTGCCGCGAGT	TACTTCAGGA	TAATTTCCCTT	8940
	TATAGTATGA	TTGATAAATG	TCACTGTTAA	CTGGCGCGGG	ATGTAGATAA	ATCGGTACGT	9000
	CTAAATTTTT	AGCTGTTTTG	AAAATAATGT	CATATTTGTC	TTGATCAAGA	AAACCATCTT	9060
50	GTGCACGTCC	CATAATGAGC	GCACCTTTGA	ATCCTAAATC	ATTGATGCAA	CGTTCGAATT	9120
	CTCGCGCTGC	GGCTTCAGGC	TCATTGATAG	GTAAAGTTGC	AAAGCCTACA	AAGCGATTGG	9180
55							

	TCTGACCAAC CAAATTTGAA GGAGAACCAT TTCCATAAGA TAAGACTTGA ATTTGAACGT	9300
	CTTGATTATT CATAAATTGG ATACGTTTCAT CATGATGTGA TAATTCGTCG GCATTTGTAA	9360
5	AACCTGTCTT TTTTCAAGG CCTTCTAACA TTACTTTCAT CGGTACACCT TTAGGATCTG	9420
	CTGATATCGC ATTCATCGTT TCTTTTTGAA TATCTTCAAT GACATAATGT TCTTCAAACG	9480
10	TAATACTTTT CATTTACTTC GCCTCCATAT TGTATTGCAT GTTTATTGCA TCTATTGCAG	9540
	AAGCATTTTT TATATACCTC TAATTTCAAT GTTTGTAAAC TAAAATTGAT CTACCAAGGC	9600
	ATCTCTCCAT CGCCATTAAT AAATGTACCT GTTGGGCCAT CTGCACCAAT CGTTGCTAAT	9660
15	TGAATGATTG GCTTGATTCC TTCAGAAACG TGTTTGAAT TATTACTAAA ATCACCAACT	9720
	AAATCAGTAT TTGTAGCGCC TGGATCAGCA GCATTGATTT GCATGTTAGG TAATCCTTTA	9780
	GCGTATTGTA GCGTTAGCAT TGTTACTGCC GATTTAGACG AACAATAAGC TAATGAATTC	9840
20	ACTTTAGATT CAGCTGTTTC GGGGTTTGTA ACCATTCCAA ATGAACCTAA ACCACTTGAT	9900
	ACGTTGACGA CAACAGGTTG TTCAGATTTT TCTAAGAGAG GGACGAATGT ATTCATCATT	9960
25	CGTACGATAC CGAATACATT CGTTTGATAT ACTTCTTCAA CGTCACGAGG TGTCAATTTG	10020
	GAAGGTGCTG AAAATTGACC AGATATACCT GCATTGTTAA TGAGGATATC AAGACGGCCT	10080
	TCTTTTTCAG CAATCATGTT ATAAGCATTT TTGACTGAGT AGTCACTTGT AACATCTAAT	10140
30	TGTACATAAT GAACACCTAA TTTTGTGAT GCTTGTTGTC CTCTTACATC ATTCCGAGAA	10200
	CCTATATAAA CTTTGTAAAC CAATGCTTTA AGTGCCTCTG CACTTGCATA GCCTAACCTT	10260
	TTATTGCCTC CTGTGATTAA CACAATTTTA GTCATTACGT CCCACCTCAT CTAAATAAAT	10320
35	GTTTAATAAA TAATTTCTGT ACGCTTCAAT TGAAATATGG CGATGCTCTA TTTGGAAGGC	10380
	AAATACACTA GTTGATAATG ATTGCAACAG CATATCTGTT TTGAATTCGT GTAAGTGTG	10440
	TCATCGCTTT TAAATAAGTC ATAATAAAAA TCAAATAATT CTTGATAAAA TGCGCTTTGG	10500
40	TAAAAACGTA ATTTATTGTT GCCTGCTTCA ATACATTGCA GTAGTGCCTT ATTATCGATT	10560
	TTAAATTGTA AAAGATAATC TAACGACACT TGCATAACCT CATAATTAGA ATGATAGTCA	10620
45	TCTTTAATTT GCTTAAAATG AGTGATAAAA ATATCAAGGT CTCTTTGTAT GACGTAGTAG	10680
	CATAAATCGC TTTTATCTTT GAAATGTCCA TACAATGTCC CCATACCGAT ACCTAGTTCT	10740
	TTAGCAATAC GATTCACT AATGTTTTCA ACGCCTTCTT CATCAAAAAG TTTGTGCGCT	10800
50	ATTTCTTCAA TTCGTTGCCT ATTCTCTTTT GCATCTTTTC GCATGATTAC ACCTACTTAA	10860
	AATTCTCTAA AATTGACAAA CGGATAACTC TCCGTTTATT ATAAAACGTG TTAAGAAAGT	10920
55	TAGCAATGAA TTTGCAATAA CTATTAAATA TCATAAAAGA AAAGAGTGTT GATAATGTCT	10980

ACCTTATCGG TTCAAATGAT TGCTGAAAAA CTGAATGTCA CTACAGAAGA TGTGGAAAAA 11100  
 GTATTAGCTA TGACAGCGCC ACTAGGCATT TTTAGTCATC AATTACAACG ATTTATTCAT 11160  
 5 TTAGTATGGG ATGTCAGAGA TGTAATAAAC GACAATATTA AAGGAAATGG ACAAACACCA 11220  
 GAACCATATA CGTATTTTAA AGGTGAAAAA GAGGACTATT GGTTTTTAAG A 11271

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6261 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 15 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

20 CAACCCGTTT AGAACAAAAT AAAAACCGTA CAATTTTATC ATCTTAATGA TTATTGTACC 60  
 GAAAACTTT TTTACATCAT ATCTGCATGT GCATAATCGA TATCGGTAAA TTTATTATAT 120  
 TGTTCATAA AATGTAACCT AACTGTGCCT GTTGGACCGT TACGTTGCTT AGCAATGATA 180  
 25 ATTTCAATTT CACCGTTTTT ATCATTCTGT TGTGGCTCGA AACCACCATC ATCGTCATCA 240  
 TCTTCATCGC CGCCACGGTT ATAGTAATCA TCACGGTATA AGAATGCAAC GATATCGGCA 300  
 TCTTGCTCAA TCGAACCAGA TTCACGAATA TCACTCATCA TTGGACGTTT ATCTTGTCGT 360  
 30 TGTTCAACAC CACGAGATAA CTGACTTAAT GCGATAACTG GACATTTTAA TTCACGGGCT 420  
 AATGCTTTTA ATGTACGAGA GATTTTCAGAA ACTTCCTGTT GTCTGTTATC GGACGCACGT 480  
 35 GAACCACTAC CTTGAATCAA CTGTAAGTAG TCAATCACAA TCATGTCTAA GCCATGTTCT 540  
 TGCTTTAATC GACGACATTT AGAACGTAAA TCATTAATTC GAATACCCGG TGTATCATCA 600  
 ATAAAAATCT TCGTACGTGA TAATTTACCT ACCGCTATAG TAAAACGACT CCAATCTTCC 660  
 40 TCAGTCATAG TACCCGTTCT TAAGCGGTTT GAGTCAACAT TTCCAGAACT ACAAATCATA 720  
 CGTGTGGCTA ACTGATCAGC ACCCATCTCT AGCGAGAAAA TACCAACTGT ATACATATCT 780  
 TCATGCGTTG CAACTTTTTG TGCAATATTA AGTGCGAACG CAGTCTTACC TACAGATGGA 840  
 45 CGCGCTGCAA GGATAATTAA ATCATTTTCG TTGAACCCTG CTGTCAATTTG GTCTAAATCT 900  
 CGATATCCTG TAGGTATACC TGGTGTGTTGA CCACTATTTT GATCAAGCTC TTCAGCTGTT 960  
 50 TCATACACTT GTCCTAAGAC GTCTCGAATG TCTTTAAAGC CATCGCTTTC ACGAGAAGAT 1020  
 GATAGCTCTA AAATTCGACG TTCTGCATCA CTTAAATCG CATCTAGTTC AAGTTCATCA 1080  
 TTATATCCAT CATTGGCAAT ACTATCTGCA GTTTGAATCA ATCTACGTTT TAATGCATGC 1140



	TCTGCAAGAT ATTGCGGGCC ACCCGCTTcA TTCAACGTAC CTTCCGTCGA TAATTGATCC	1260
	ATCAATGTTA CAACATCAAT TTCTTTATTA TCTTCATTTA AGTGCATCAT TGCACGGAAA	1320
5	ATATGTTGAT GGGCACCCCT ATAAAACGAC TCAGGAAGCA AAACCTCCTG AGTAGTATTA	1380
	ATCAATTCTG GATCTATAAT AATTGAACCT AAGACAGACT GTTCAGCTTC ATTGTTATGC	1440
10	GGCATTGAT TTTGCTCATA CATTCTATCC ATGAATGGTT ACACCTCTTA TTTCAATCCA	1500
	ACTTTATTGT TCAACTGTGT GTACGCGAAT TGTACCTTCA ACTTCTTTAT CTAATTTAAC	1560
	AGGTACATTC GTATATCCTA GGGAAATGAAT TCCATTGGT AAATCCATTT TACGTTTATC	1620
15	AATTTTAATA TCATGTTGTG CTTTtagTGC TTCGGCAATT TGTTTTGTAC TTAGTGACCC	1680
	AAACAATTTA CCACCTTCAC CAGTTTTTGC TGaTACTTCA ACTTCAATGT TTGATAACGT	1740
	TTCTTTTAAT GCTTTAgCAT CTTCAATTTT TTGTTGGCGT TCTTGTTTTG CACGTTTTTT	1800
20	CTGTAACCTT AATTGTTTAA GGTACCTGG TGTGCTTCT ACAGCATAAT TCTTTTTCAA	1860
	TAAGAAGTTA TTTGCATAAC CTACTGGTAC TTCTTTAACT TCACCTTTTT TACCTTTACC	1920
	TTTACCTTTA ACATCTTGTG TAAAAATTAC TTTCATGCAT CTTCACTCCT ACTTAATTGT	1980
25	TCTGTAATTG CTTGTTGTAA TTGTGCTATC GCCTCTTCGA CTGTACACACC TTTAAGTTGT	2040
	GTTGCCGCAT TGGTTAAATG TCCACCGCCA CCAAGTGCTT CCATTGTTAA CTGGACATTT	2100
30	ACTGAACCGA GTGAACGCGC AGATATACCA ATCAGATTAT CTTCACTCTT CGCAACAACA	2160
	TATGATGCTT CAATACCTTC TAAACTTAAC AGTTCATCTG CTGCTTGTGC AACTGTTACT	2220
	GGATGATAAA TTTTATCGTC TGAACCATGC GcAATGGCTA TGCCATTATC TTCAACTTTT	2280
35	ACAGTTCGAA TTAATTCAGA TCGATTAATG TAAGTATCCA CATCATCTTT TAAGAAATGT	2340
	TGCGTTAAAA TCGTATCTGC ACCATGTGCA CGTAAATAAC TCGCTGCATC GAATGTTCTT	2400
	GATCCTGTTC GTAATGTAAA GTTTCTTGTA TCTACAATAA TACCTGCATA CATCACTGTT	2460
40	GATTCAAGAC GTGTTAAACG TTGTTCTGTT GGTGATATT CCAGTAACTC TGTTACCAAT	2520
	TCAGCTGTCG AACTTGCGTA TGGTTCCATA TATATCAACA ATGGATTAGA GATGAAGCTT	2580
45	TCACCACGTC TATGATGATC GATAACAAC TTTACGGTTTG CTTTATTTAA GACATTTTCA	2640
	TCTAAAACCA GTTCCGGTTT ATGCGTATCA ACAATCACTA CGGTTGTCTT AGATGTCATC	2700
	ATATCCCAAG CATCATCTGA TGTAATAAAT CGCTCTCTTA ACTCTGGCTT TTTATCTATT	2760
50	TCGTTTCATCA CGCGTCGTAA TGTTGGATCA ATGTCAGTCT CATTTAATAC GATGTATGCT	2820
	TCTAAATTAT TCATCATTGC AAATCTAGAC ACACCGATTG CTGCACCAAT TGCATCTAAG	2880
55	TCAGGACGTT TATGTCCCAT GATAATGACT TTGTCACCCT CTGCAAGGAT ATCTTTTAAC	2940

CCATAGAAAC GCACATTACC ATTAATACTT TTAATTGCAA CTTGGTCGCC ACCGCGTCCT 3060  
 AATGCTAAGT CTAGGCCTGA TTGTGATAAT TCACCTAAGT CGATTAAATT TTCAGTACCT 3120  
 5 TCACCAACAC CGATACTTAA TGTTAATTGG GCACGATAAC CAACACTTTT TTCACGTAAT 3180  
 TGA CTCAAGA TATCAAATTT AGATTCTTCT AAGTCAGCTA ATATTTTTTG ATTAAATAG 3240  
 10 GCTACGAATT GATCGGAACT GTATCTTTTG AAAAATATAT TATACTCAGT TGCCCATCGA 3300  
 CTAATGACAC GCGTTACCAT TGAGTTGATT TCCGAACGCT GCGTATCATT CATATTTTGC 3360  
 GTAATCTCAT CGTAGTTATC TAAAAATAAT GTCGCAATGA TTGGTTTAGA ATTTTCATAT 3420  
 15 AGTTCATTTG TTTGTACTTG TTCAGTTATA TCAAAGAAAT AGAGGCAGTG ATCATTCTCA 3480  
 GAATAACGTA CTTGGAAATG ATACTGATTA TATTCTATTT CAACGGATTT CACTCTATCT 3540  
 AATTGCTTTA AAATGTTTGG AAATACTTCA TTTACAGATT CAGAAATGAC ATTCGCTTCC 3600  
 20 ATATGATCTG TCATAAATTG GTTAACCCAT TCGATGTGAT CATTTTTCATC TAAAAAATG 3660  
 ATACCAATTG GTAAATGTTT GATTGCTTTA TTATTTGTTG TTGAAATTTG AGCACTCAAA 3720  
 CCATCTACAT AACTATCCAT TTTCAATAAA GCTTGTCTGA ATAAAATGAT GCTAACAATA 3780  
 25 ATCATCACGA CAAGAACGAT AGATGCAATT AGTGCTATAA GACTATTAAA GATAAACCAT 3840  
 ACACCCATTA AAACAATTGC TGTGATGATC ATGATGACAA ATGGTATTAG TAAAGCTTTC 3900  
 TTAGTGGACT GCCGATTCAT TATTCCACCT CTATTCACCT TTTAGAATTA TTTTTCATGA 3960  
 30 TTCGCTTCAA ATTCAAACCT AAATCGATAA CACCAAGTAG TCCTACAATA TGTGTCGTAG 4020  
 GTGTCAGTAT TGTACCGATA ACCAATAGTA AAATCGTTAC TGCATTCCGC AAACCTTTTCG 4080  
 35 CTTTACCAAA GAAATGAATA ACACTTAAAC CTGAATATA CATTACTAAT GATAACACAA 4140  
 GTTGGAAGTT TAAAAGAATG CTCTGGAACA CACTCGGTTG ACCTGTAAAT AATAAACATA 4200  
 TGATAACAAT AATGTATATC CATAATAAAA TACCGTCAT TTGCCACGCG AAAAGTGGCT 4260  
 40 TAAATACAGG TGTAGCGATT TTAAATTTTC GTAAAATCGG AAATGTAACG ATTAAGTTAA 4320  
 TTAAGACGAT TAAAAATGTA ATGATAATGA TGAAACCTGG TAATTGAACG GTCGCTTGTC 4380  
 TAAACCCCTC TTCTAATATT TGGGTCATAT TCGCATCGGC ACCGCTCATC GTAATCGCTT 4440  
 45 CATGTAATGT TTGCTTGAAA GGTTTTACTA TGCTCGCTGA TGGTGGAATC CTTCCGAATG 4500  
 TTTGTAGTAA CATAAAAGCG ATTAATGAAA TTAACTCAT CGCTACTGTT GTTACGTATA 4560  
 50 ATATTCTTTC TTAGACGTT CTTTCTTTGA GCAATTGACC AATAATTAAA CTTGCAATTA 4620  
 AGACTAATAT GATGGCACTT AAAACGAAAG TATTACCTAA AACAGTTGTT ATAATTACTG 4680  
 TAATAAGTGC ACTAATCCCG AAAGATTGTA TTGATTTATT CCATAAACG ATACCTGGTA 4740

55

CAAATACCAA CGCAATCGTT GCAATTATTG TTGCTTTAGG TTGTATTTTT GAAAACACAT 4860  
 AAGCCACTCC CATATTTTAA ACTATAGCTA TTATTTTAAC CTCTTTAATG AAAATTAACA 4920  
 5 ATTTATAGAT TGTATGCTTC TATTTCAATTT AATTGAATAA TAACTTTCAT GTTTTATAAG 4980  
 TAATTAACAT ACTCATTGGA ATCGCTTTTG TGTGCTTTCA TTTTCAACAT GATTATTTAA 5040  
 TCCCACTACA TAGCAATCAA GCTTGATTTA GATTTACAAT ACATTTCCAC TCTCATGTAC 5100  
 10 TCTAGATGTT TTTGAATATG ATAAGTGTGA TTTAGTGGCT TCATTCTTTG AAAATATATA 5160  
 TTATTACTTA CGCTTAAAT GCTTTAAATT TAAGAAATGA TATAAGTTAG GTGCCCAGGT 5220  
 ACTAAAGTTT AGTAGGATC CATCATGCCC AACATTATCA GGCACGAAGA AATGACGATG 5280  
 15 ATATTTAAAA CGTTCACCTA ATGCACGAAC TTGATCATCC GGATATAGCA AATCATCTAT 5340  
 GAACCCCATC GTTAACACTT TTGTTTCTAA ATTTTAAAA ACATGCGTTA CGTCTGTGCG 5400  
 20 ACCTCGGTCA ATGTTGTGAC TATCCAATAC ATCTAGCAGT GTCAGATAAC AATTCAAATC 5460  
 AAAATGTTCT TTAAATTTAT TACCTTGATG TTGTTGGTAT GCGACTACTT CATCCGGCGT 5520  
 AAAACGTTCA TCATAACTTT TTGATGATCG ATATGTCAAA AAACCTAATT GGCGTGCAAT 5580  
 25 ACTTAGACCT TCCTTACCAC CAAGATGAAT GGCTTGCTT GCAATTTTCAT TGAAAGCTCT 5640  
 ACTATAAGAT GATGTTGAC TTGTTGCAGC AAGGATAATG GCTTTATCTA CTTCAAACCTG 5700  
 TTGATTGTAG AGTAGTTCCA TTGCTTGCAT ACCTCCAAGA CTCCCCCTA TTAAAAATATT 5760  
 30 AATCTTATCA TAACCAAGGG CTTGTATACC TCGTTCATTC GCTCTGACTA TATCTCTTAA 5820  
 TGTTAATTTT TTAGGAAAAT GAGGGTCGTT TAAAGGTGAA CTGAACCGA AAGGACTACC 5880  
 AATAACATCA AATGTTAAAA ATTGATAATC GTGAATGGGT ATATATCCCC CATCAATAAT 5940  
 35 TTCTCGCCAC CAACCCGGAT AATCATCTGT TCCATATGTT AAATGATTGC CAGTTAATGC 6000  
 ATGACAAACT ACAACTAATG GTTGTCCATG ATAACCGACA TGCTCATATC TCAAACGCAA 6060  
 40 GTATCTATG ACTTCCCCAG ATTCTGTAAT AAATCCCCT AAATTTAAAG TATCTACTGT 6120  
 GTAATTTGTC ATTGTTCTTT CCTCCTTAAA CAAAAAACT TCTCACCCTA TTGAAAAGTA 6180  
 AGAAGTCTTT ATACTTATCA TTCGAGTAAC TCGTTGGTTT TAGCACCGTG CTATAAAGTC 6240  
 45 GGTTGCTGAA GTATCACAGG G 6261

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

- 50 (A) LENGTH: 1222 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

55

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

5 ATGCGATTAA CTCTGGAAAT ATCTTTTCCA TATTTACGTh TTAAATTATT CAGCAAATTC 60  
 ATACGAGaTT CATACTCGTT yAACACTTGT TCGTCGAATT CTGTATTAGC CATTTTCATCA 120  
 TATAACTCAT GTTTTGCATC TTCTAAAATG TAGTAAAATT GATCAATATC TTCTTTTAAT 180  
 10 TTGTCATAIT TGTTTGGAAC TATATCGTTT ATTGTTAACA AATGGTTGCT TAGTTCATAT 240  
 AAACGATCAG TGATAGCATT TTCATCCGTT AATGTCATAT ATGCGTTATT AAGCGCTAAG 300  
 CTTAATTTTT CAGAGTTTTG AATGCGTTTA ATATCTATTT CAAGTTGCTC TATTTGCGCT 360  
 15 TCTTTTAGAT GTGCTTCAGA CAATTCTTCT AATTGGAATT TCATTAAATC TAAACGCTGT 420  
 AGCAATGCTT GGTCTGCTGA TTCTAAATCT TCTAACTCTT GCTTTTGGC TTTATAATTT 480  
 TGAAAAGTTT GGTGATATTT ATCCAACAAA TCTTGATAAC GTGATTCTGC GTAATTATCC 540  
 20 AATAATGTTA AATGGTATTT TTGTTTCAAC AAAGACTGCG TTTTCATGTTG GCCATGAATA 600  
 TCTAATAATT CTTGCATAAC TTTTCGTAAA TCTTGTAAG TAACTGTTTG ATTATTAATT 660  
 TTACAAAGAC TTTTACCAGA GCTGAAAATT TCCCGTTTAA CTAATAAAAA ATCTTCATCT 720  
 25 ACATCAATAT CCATATTTTT CAATATATGT ATAGCATCTT TACTCTCGTC AATATCAAAT 780  
 ATACCTTCGA TGACAGCCTT TTTTCCACCA TGTCTTACAA AATCAGATGA AGCTCTCATT 840  
 CCAATTAATT GTCCAATTGC ATCTATAATA ATTGACTTAC CTGAACCCGT TTCACCACTT 900  
 30 AAAACAGTTA AACCATCAGA AAATTGAATT TCTAATTCTT CAATAATAGC AAATTGCTTG 960  
 ATTGATAAGG TTTGTAACAT AAATCATCG CATCCTTATA ACAAATTGAA AATTCTTGAC 1020  
 35 TTGATTTTCAT CACTTGCCTC TTTGCTTCGA CAAATAATTA AACAAGTATC ATCACCACAA 1080  
 ATTGTGCCTA GTACTTCTTC CCAATTGATT TGGTCTAATA TAGCTCCAAT AGATTGTGCA 1140  
 TTACaAGGTA TGTTTTTAGA ACAAGTAAAT TATCAGTACC ATCTATATTA ACAAAGGAAT 1200  
 40 CCATTAAATA ACGTCCCAAT TT 1222

## (2) INFORMATION FOR SEQ ID NO: 14:

## (i) SEQUENCE CHARACTERISTICS:

- 45 (A) LENGTH: 1021 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

50 TTTGTTATTA TTACnThAAA TAATTGCATT ACTTTTTACT GATGGTACAA CTTTCCATCC 60  
 55

TTCTTTTGGC ACGACATAAT TGTCTTTATC TTGAAGTAAA TATCCGCCAG ATACTGAAAC 180  
 AAACCTCTTCT TCGTTACTGT CTATAGTCAT ATCAATTTCT AATAATCTTA CATTCTTCTT 240  
 5 TTGTTTTTAAA ATATCTAATG CTTTCATCTGT AAATTTTGGT GCAATAATGA CTTCCAAAAA 300  
 GATACTATGC AATTGCTCTG CTAAGTCAGG TGTTACAGCT CGGTTTAATG CAACAATTCC 360  
 ACCAAATATT GATTGACTAT CCGCTTCATA CGCATGTTGA AATGCTTGTT CTATCGTGTC 420  
 10 ACCGATACCA ACACCACATG GATTCATGTG TTTAACCGCA ACTGTAGCAG GTGTATCAAA 480  
 CTTTTTAACT AAAGCTAGTG TAGCATCTGC ATCTTTAATA TTGTTATAGC TTAATTGTTT 540  
 CCCATGTAAT TGTTTAGCGC CTGCAATCGT GTGCTTAGCA TTCGAAGTTC TCACAAAATA 600  
 CGCTGATTGT TGTGGATTTT CTCCATATCT TAAAGTTTCT TTATCCCCTT TAAAGAAAACG 660  
 TACAATCGCT TCATCATATT CTGCAGTATG CTCAAAACT TTAATCATTG ATGATTGTCT 720  
 20 ATATGACTCA TCTAACGAAT CGTTTCTTAA TCGCGTCAAT ACTTCTTGAT AATCTGCCGG 780  
 ATGTACAATT GTTGTTACAT GTTTATAGTT TTTAGCTGCA GCACGTAACA TTGTTGGACC 840  
 ACCAATATCA ATATTTTCAA TTGCTTCGTC CATCGTCACA TCAGGGTTTG CAACAGTTTG 900  
 25 TTGGAATGGA TATAAATTAA CTACTACCAT ATCAATTAAA TCTATATGTT GTTCTGATAA 960  
 TTCATTTAAA TGCTGCGGTT TATTTGATC AGCTAAAATG CCACCATGAA CAGCCGGATG 1020  
 T 1021

## (2) INFORMATION FOR SEQ ID NO: 15:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3759 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

TCATTCACTC CTAAATTGTT ATTACACTAT TACACaTAGC TAATCATCAA TGTGAAATCA 60  
 CCTTCAAAGA CACTATCCAA ATCTTCAGAA GTCAAAATAA AGTTTGTACC AGTAGTCACT 120  
 45 TTGAAAATTT CACCATCGAC AATCATTTGC CCTTCGCCTT CCAACACTGT AACTAAACAG 180  
 AACTCTCTAG GCTTCATATA ATTTAACGTG CCAGAAATTT CCCATTTAAC CAATGTAAAG 240  
 AAATCATTCTG ATACAATGTG TGTACACTTA TGGTTTTCAA TAATTTGCTT TTCAGGCAAA 300  
 50 ATATTAGGTA ATGGTGCATT GTAGTGAATA ACGTCTAAAG CTTTTTCAAT ATTTAACGGT 360  
 CTATCATTAT ATTGATTATC TTGACGATTG AAATCATAAA GTCTATATGT AATGTCTGAC 420

ATAAAAtAGa ATTcYCCAGG kTTTACtTTA AtatATCyAA gTAtCGaCtC tATCGTTCCG 540  
 TGTTGAACAT GATTGCGAAC TTCTTCTCTA GACTCTGCTA ATGTCCcTAT AACTATTTCT 600  
 5 GCATCTTCTT CTGCATCTAT AATATACCAa CATTcAGATT TGCCATATTG CCCgTTTTCA 660  
 TGCTCATAAG CATAAGAATT ATCAGGGTGC ACATGAATAG AAAGTGATTc TCTTGcATCC 720  
 ACTATTTTAg TTAGAAGCGG AAAATCTTTG CTTGGGAAAT CACCAAAcAA TTCACGATGT 780  
 10 TCTGACCAAA TACGGTCTAA TGTTTGACCT TGATATGGTC CATTaATAAT CTCGCTCGTA 840  
 CCATTTGGAT GTGCTGACAC ACACCAACAT TCCCCcAGTT GTATcATTGT CTAATTGATA 900  
 15 TCCAAACTCA CTTAGACGTT GACCGCCCCA TAATTTTGTT TTTAAaATTG GTTGTAaaaa 960  
 TAATGGCATT GTTGcACCTC CATTGTGATT AAGTAAGCAa TAGAACTCTG ATGTTGTTGT 1020  
 TCCATTATAT TTTGATTTTG TTCTcATTa CATCGTATTA TTAActTCCA CATTTCaAAT 1080  
 20 TAACTATTAG TGATTGTACC ATATTTACTA ACATTGcAGT ACTGCCAATT AAAAGnGCTT 1140  
 CACTTAAATT TACAGTACTT TAACATTTTC AAAAATTaT AGCATAGAGA TTATATCTCT 1200  
 CTTACATTTG TACATATTTc CCTTTAAATT TACTCGCCCA TTATACCAAT TAATaAaCAa 1260  
 25 CTTTAATAGT TGTGCCATAC ATTGTTCAAA TTCTTTGTAA AACGCATAGA CAATACGTAC 1320  
 TTATTCATAC TTATAATTCA TCATTTTCaA AAAATAACGA GTTACGAAAA AGTAACCCGC 1380  
 TTCAaATCAT ATTTACTATC CTTATTaATC CGTTTCATTT TCAaATTGAG TTAAAGCATC 1440  
 30 TTTAATGTCC TGATCACCAC TAATAATTG AAACTCTTGG TGATTAAaAT GATTGGATGT 1500  
 GACAATTTCT TTTAATACTG TCGCAACATC TTCTCTAGGA ATTTcACCTT TACCATCAAA 1560  
 35 ATATTGTGCA GCTTCTATCT TTCCAGATCC TGCTGCATTT GTAAGTGCCC CTGGATGTAA 1620  
 AATTGTATAA TTCAaACCTG nAACGTCTTA AATAGTCATC AGCGTAATGT TTAGCTATTG 1680  
 TATATGGCTT TAAATCACCG CTATCATCAa AAGCCTGACG TCTCGAATCA TATGTTGAAA 1740  
 40 CCATGACATA GTGTTTAATA TTGGCCTCTT TACTCGCAAT CATTGATTa ACAGCACCAT 1800  
 CTAAATCGAC AATAATTGTT TTATCTGCAC CCGTGTCCC TCCAGAACCT ACTGAAAAGA 1860  
 TAACTTTATC GAATGGTTTA AACGTCTCAG TTAAAGTCTC TATTGAATCA TTTTCaACAT 1920  
 45 CAACAAGAAT TGCTTTcATA CCTTGtGATT TTAACGCATT AAGTTGATCT GATTGCCTAA 1980  
 CACCAGCAGT AAATGGTACA TTTTCTTTTG CTAATTGTTG CACTAGTAAC GAACCTACAC 2040  
 CGCCATTAGC ACCTATAACC AAAATATTCA TTTACAACAC TCTCCTATkT ATTATTCTCT 2100  
 50 ATGCCATACC ACTTTATGAG ATATGTAAAA CTTGTTACAA CTATAAAaAT CAATTGACAT 2160  
 ACTACTGGGA ACGTATTAAA TTAATATATG AACAAATATT CATATGAAAG GATTGTcATA 2220  
 55

tCaAGGCATT AGcGATTACA ATCGAATACG TATCaTGGAA TTGTTATCaG TCAGCGAAgC 2340  
 AAGTGTGGT CACATTcCAC ATCAATTGAA TTTATCTCAA TCAAATGTCT CGCACCAATT 2400  
 5 AAAATTACTT AAAAGTGTGC ATCTTGTGAA AGCAAAACGA CAAGGCCAAT CAATGATTTA 2460  
 TTCATTAGAT GACATCCACG TAGCAACTAT GTTAAAGCAA GCCATACATC ACGCGAATCA 2520  
 10 TCCTAAAGAA AGTGGGTAT AATATGTCTC ATTCACATCA TCATCATGAC CATATGCATA 2580  
 GTCATGTAAC TACAAATAAT AAGAAAGTAT TGTTTATATC GTTTTTAATA ATCGGTCTAT 2640  
 ATATGTTTAT CGAAATCATC GCGGTCTCC TTGCTAACAG CTTGGCATT CTATCTGACG 2700  
 15 GTATCCATAT GTTTAGCGAC ACATTCTCAT TAGGTGTGTC ACTTGTCGCA TTTATTTATG 2760  
 CTGAAAAGAA TGCCACAAC ACAAAAACAT TTGGTTATAA ACGTTTCGAA GTACTCGCAG 2820  
 CGTTATTTAA CGGTGTAACG CTTTTGTAA TAAGTATTTT GATTGTTTTT GAAGCGATTA 2880  
 20 AACGTTTCTT TGTTCTTCT GAAGTTCAAT CAAAAGAAAT GTTAATCATT AGTATTATCG 2940  
 GTTTAATTGT CAATATCGTT GTTGCAATTCT TTATGTTTAA AGGCGGCGAC ACTTCACACA 3000  
 ATTTAAATAT GCGTGGTGCT TTTCTACATG TTATCGGAGA CTTATTAGGT TCAGTTGGCG 3060  
 25 CCATTACTGC AGCTAkTTTA ATTTGGGCAT TTGGATGGAC AATCGCCGAT CCTATCGCAA 3120  
 GTATTTTAGT TTCCGTTATT ATTTTAAAAA GTGCTTGGGG TATCACAAAA TCTTCAATTA 3180  
 ACATTTTAAAT GGAAGGCACA CCAAGTGATG TTGATATAGA TGAAGTTATA ACTACTATTA 3240  
 30 AAAAGGATTC ACGAATACAA AGTGTGCATG ATTGCCATGT TTGGACAATT TCAAATGATA 3300  
 TGAATGCATT GAGTTGTCAT GTTGTGTAG ACCATACATT GACAATGAAA GAATGTGAAT 3360  
 TATTATTAGA AAaCATTGAG CATGATTTAT TACATTTAAA TATTCACCAT ATGACTATTC 3420  
 35 AATTAGAAAC GCCTAATCAC AAACATGATG AATCGATTAT ATGTTCAGGA ACACATAGTC 3480  
 ATTCACATAA CCATCATGCT CATCATCAGC CGCATGTACA TTAATAATTT TAACCTACTG 3540  
 40 CCATTGCATC GATTAAACTT TTCAATGGCA GTAGGTTTTT TATGCTTTA TGGCGACTTG 3600  
 TTTGGTCTTT GATGATGCAA TGTTTATTAA CAAATTTTCA ACTATTATTT CTTACATTAG 3660  
 TCATATTTT GACAATTTAC TATTATAATT CTCTAACTTT AGTCACTTTA ATTAATTTTT 3720  
 45 ATTAGATATT AATATGAAAA TAACGTGTTT TTTGTTATT 3759

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13086 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

	TAATTATCGC GCATAACAAA ACATTAGCAG GACAATTATA TAGTGAGTTT AAAGAATTTT	60
5	TTCCTGAAAA CAGGGTGGAA TACTTTGTAA GTtACTATGA TTATTATCAh CCAGAGGCAT	120
	ACGTACCGTC TACTGACACT TTTATTGAAA nAGATGCCTC AATCAhTGAT GAAATTGATC	180
	AACTACGACA TTCTGCTACA AGTGCATTAT TTGAACGCGA TGATGTAATT ATTATTGCTA	240
10	GTGTAAGTTG TATATATGGT TTAGGTAATC CTGAAGAATA TAAAGATTTA GTAGTAAGTG	300
	TTCGAGTTGG TATGGAAATG GATAGAAGTG AATTACTTAG AAAACTTGTc AGATGTGCAA	360
	TATACACGAA ATGACATCgA TTTcCAACGA GGAACGTTTC GAGTGCGTGG TGATGTAGTG	420
15	GAAATATTCC CAGCCTCTAA AGAAGAACTT TGTATAAGGG TTGAGTTTTT CGGCGATGAG	480
	ATTGACCGTA TCCGAGAAGT TAACTACCTA ACAGGTGAAG TGTTGAAAGA AAGAGAACAT	540
20	TTTGCGATAT TCCCAGCTTC TCACTTCGTA ACACGTGAAG AAAAGTTGAA AGTTGCGATT	600
	GAACGTATTG AAAAGAATT GGAAGAACGA TTGAAAGAAT TACGAGATGA GAATAAATTA	660
	CTAGAAGCGC AAAGGTTAGA ACAGCGTACC AACTATGATT TAGAAATGAT GCGAGAGATG	720
25	GGATTCTGTT CAGGAATTGA AAATATTCC GTACATTTAA CTTTGCGACC ACTGGGTTCC	780
	ACACCATATA CTTTATTGGA TTACTTTGGC GATGATTGGT TAGTAATGAT TGATGAATCA	840
	CATGTGACAT TACCGCAAGT TCGAGGCATG TATAACGGAG ACAGAGCGCG TAAACAAGTT	900
30	TTGGTGGATC ATGGGTTTAG ATTACCGAGT GCATTAGATA ACCGTCCACT TAAATTTGAA	960
	GAATTTGAAG mAAAGACAAA ACAACTTGTG TATGTATCTG CAACGCCTGG ACCATACGAA	1020
	ATTGAACATA CGGATAAGAT GGTGAACAA ATTATTCGTC CTACTGGTTT ACTGGATCCT	1080
35	AAGATTGAGG TTAGACCTAC TGAAAATCAA ATTGACGATT TATTAAGTGA AATTCAAACA	1140
	AGAGTgAGCG TAATGAACGC GTACTTGTTA CAACGCTCAC TAAAAAGATG AGTGAAGATT	1200
	aACCACATAC ATGAAAGaAg CGGGTATTAA aGTtAATTAT CTGCATTcAG AAATCAAGAC	1260
40	ATTAGAACGA ATTGAAATAA TTAGAGACTT ACGAATGGGT ACATATGATG TTATCGTAGG	1320
	TATTAATTTA TTAAGAGAGG GTATTGATAT ACCAGAAGTT TCTCTAGTTG TCATATTAGA	1380
45	TGCAGATAAA GAAGGGTTTT TACGTTCTAA CCGTCATTA ATTCAAaCAA TAGGTAGAgC	1440
	TGCGCGTAAC GATAAaGGTG AAGTCATTAT GTATGCCGAT AAAATGACTG ATTcGATGAA	1500
	GTATGCAATT GATGAGACAC AACGTCGTCG AGAAATACAG ATGAAACATA ATGAAAAACA	1560
50	TGGTATTACA CCTAAAACAA TTAATAAAAA AATACATGAT TTAATTAGTG CTACTGTTGA	1620
	AAATGACGAA AATAATGACA AAGCACAAAC TGTGATACCT AAGAAGATGA CGAAAAAAGA	1680

55



	TTTCGAGAAA GCTACAGAAT TAAGAGATAT GTTATTGAA TAAAAAGCAG AAGGGTGACA	1800
	AGTAAATGAA AGAACCATCC ATAGTAGTAA AAGGTGCTCG TGCGCATAAC TTGAAAGATA	1860
5	TTGATATCGA ACTACCTAAA AaTAAATTAA TTGTTATGAC AGGTTTATCT GGGTCAGGTA	1920
	AATCGTCATT AGCATTTCGAT ACTATATATG CTGAAGGACA ACGACGTTAT GTTGAATCAT	1980
	TAAGTGCCTA TGCGCGTCAA TTTTITAGGCC AAATGGACAA ACCAGATGTT GATACAATTG	2040
10	AAGGATTATC GCCAGCAATT TCAATAGATC AAAAAACAAC AAGTAAAAAT CCAAGATCAA	2100
	CTGTAGCAAC AGTAACAGAA ATATATGATT ATATACGTTT GTTATATGCA CGTGTTGGTA	2160
	AACCTTACTG TCCAAATCAC AATATAGAAA TTGAATCGCA AACAGTACAA CAAATGGTTG	2220
15	ACCGCATTAT GGAATTAGAG GCACGTACAA AGATTCAATT ATTAGCACCT GTCATCGCTC	2280
	ATCGTAAAGG TAGTCATGAA AAGCTAATCG AAGATATTGG TAAAAAAGGT TATGTACGTT	2340
20	TAAGAATCGA TGGCGAAATT GTTGATGTAA ATGATGTACC TACTTTAGAT AAGAACAAGA	2400
	ATCATACAAT AGAAGTTGTT GTAGACCGAT TAGTTGTTAA AGATGGAATT GAAACACGAC	2460
	TAGCTGACTC TATAGAAACT GCCTTAGAGC TTTCAGAAGG ACAATTAACA GTCGATGTCA	2520
25	TTGACGGGGA AGACCTTAAG TTTTCAGAAA GCCATGCTTG TCCTATATGT GGATTTTCAA	2580
	TCGGAGAGTT AGAACCAAGA ATGTTTAGCT TTAACAGTCC TTTTGGTGCT TGTCCGACAT	2640
	GTGATGGCTT AGGCCAAAAG TTAACAGTCG ATGTAGACTT GGTGTGTTCC GACAAAGATA	2700
30	AGACGCTAAA CGAAGGTGCA ATAGAACCTT GGATACCGAC GAGTTCTGAT TTTTATCCAA	2760
	CATTGTATAA ACGTGTGTTGT GAAGTTTATA AAATCAATAT GGATAAACCT TTTAAAAAGT	2820
	TAACAGAACG TCAACGTGAT ATTTTATTGT ATGGTTCTGG TGACAAAGAA ATTGAATTTA	2880
35	CATTTACACA ACGTCAAGGT GGTACTAGAA AACGAACAAT GGTTTTCGAG GGTGTAGTTC	2940
	CTAATATAAG TAGACGATTC CATGAATCTC CTTCAGAATA TACACGTGAA ATGATGAGTA	3000
40	AATATATGAC TGAACCTACCT TGCGAACTT GTCATGGAAA GCGATTGAGT CGTGAAGCKT	3060
	TATCTGTTTA TGTAGGTGGT TTAAATATTG GTGAAGTAGT CGAATATTCA ATCAGTCAAG	3120
	CGCTGAAC TAATAAAAAC ATTGATTTGT CAGAACAAGA TCAAGCGATT GCAAATCAAA	3180
45	TATTGAAAGA AATTATTTCC CGACTCACTT TTTTAAATAA TGTGGGACTT GAATATTTAA	3240
	CGTTAAACAG AGCTTCAGGT ACACTTTCAG GTGGTGAAGC ACAACGTATT CGATTAGCAA	3300
	CGCAAATTGG GTCGCGTTG ACTGGTGTCT TATATGTATT AGATGAGCCA TCAATTGGAC	3360
50	TGCATCAAAG AGATAATGAT CGATTAAATTA ATACACTTAA AGAAATGAGA GATTTAGGAA	3420
	ATACTTTAAT TGTAGTTGAA CACGATGATG ATACAATGCG TGCGGCTGAT TACTTAGTGG	3480
55		

	AGGTAATGAA AGATAAAAAA TCATTAACAG GACAATACTT GAGTGGTAAG AAACGTATTG	3600
	AAGTACCTGA ATATCGCAGA CCGGCTTCAG ATCGTAAAAT TTCTATACGT GGAGCTAGAA	3650
5	GCAACAATCT TAAAGGGGTT GATGTGGACA TACCACTATC AATCATGACG GTTGTTACAG	3720
	GTGTATCAGG TTCTGGTAAA AGCTCATTAG TAAATGAAST ATTATACAAA TCATTAGCTC	3780
	AAAAAATTAA TAAATCTAAA GTAAAGCCAG GATTGTACGA TAAGATTGAA GGTATTGATC	3840
10	AACTTGATAA AATTATTGAT ATTGATCAAT CACCAATAGG TAGAACGCCA CGCTCTAATC	3900
	CAGCAACATA TACTGGTGTG TTTGATGATA TACGTGATGT GTTGCGCAA ACAAATGAAG	3960
15	CTAAAATTCG AGGATATCAA AAAGGGCGTT TTAGTTTTAA TGTAAGAGGT GCACGCTGTG	4020
	AAGcTTGTAA AGGTGACGGT ATTATTAAAA TTGAAATGCA TTTTTTACCT GATGTTTATG	4080
	TTCTTGTGA AGTGTGTGAT GGTAAACGAT ATAATCGTGA GACACTAGAG GTTACTTACA	4140
20	AAGGTAAAAA TATTGCTGAC ATTTTAGAAA TGAAGCAACA CAATTTTTTG	4200
	AAAATATTCC TAAGATTAAG CGCAAGTTAC AAACACTAGT TGATGTTGGT CTTGGATACG	4260
	TCACATTAGG TCAACAAGCT ACAACGTTAT CAGGTGGTGA GGCTCAACGT GTGAaACTTG	4320
25	CATCTGAACT TCATAAACGT TCAACTGGTA AATCTATTTA TATCCTAGAT GAACCGACAA	4380
	CAGGGTTACA TGTTGACGAT ATTAGTAGAT TATTAAAGT ATTAAACCGA TTAGTTGAAA	4440
	ATGGTGATAC TGTTGTAATT ATTGAACATA ACCTAGATGT TATCAAAACA GCAGACTATA	4500
30	TTATAGACTT AGGTCCTGAA GGTGGTAGTG GCGTGGTAC TATTGTTGCG ACTGGCACAC	4560
	CCGAAGATAT TGCTCAGACA AAGTCATCAT ATACAGGAAA GTATTTAAAA GAAGTACTTG	4620
	AACGAGATAA ACAAATACT GAAGATAAAT AAGATTAAAA GAAGTGAAGG ATGTTATAAA	4680
35	TTTATCCTTC GCTTCTTTT ATTAATTTAG TAATGAATAG TAGAAAGAAA AGATGCGTAA	4740
	AAAGaATTAT GTTAAGATAG GGTCAATCTA GAGTAGTTAA ACATAAATCG AACTGGGAGT	4800
40	GGGACAGAAA TGATAAAGAA TCACTAATGA TTTATTATGT AGTGGTTCTT TGTCATTAGC	4860
	CACAGCTATT GTGTACTTAA AAATAGGaT GCaTgAGTGC AACTCATGCA TAAGaAATAC	4920
	TAATTTCTAA AGAAAAAGTA TTTCTTTATG TTGGGGCCCC GCCAACTTGC ATTGTTTGTA	4980
45	GAATTTCTTT TCGAAATTCT TTATGTTGGG GCCCCGCCAA CTTGCATTGT TTGTAGAATT	5040
	TCTTTTCGAA ATTCTTTATG TTGGGGCCCC GCCAACTAAT TCCAATATAT CATTGTAGAG	5100
	CTTAGGTCAT TGATTTTTGG CTCGGACTTT TATGCGGATA TGAACCATGT AAATTAAGCA	5160
50	AGCAATAAAT TAATGATTGA TATTGACTTG TAAAATAATA ACAATAATGA ACAATTAATA	5220
	TTTATTTTAG CTTTTCaATG TAGATTGGTG TTATATTTTT GATATGATAA GAAGAGATGT	5280
55		

	ACATTAAAGT TAGATTTAAT CGCTGGTGAA GAAGGACTAT CGAAGCCAAT TAAAAATGCT	5400
	GATATATCAA GACCGGGCTT AGAGATGGCA GGTATTTTTT CACATTATGC GTCAGATAGA	5460
5	ATACAACATAT TAGGAACAAC GGAACATATCG TTTTACAATT TATTACCAGA TAAGGATCGC	5520
	GCAGGTCGTA TGCCTAAACT ATGCAGACCA GAAACGCCTG CAATTATTGT GACACGTGGA	5580
	TTGCAGCCAC CAGAAGAATT AGTTGAAGCT GCAAAAGAAT TAAATACCCC ACTTATAGTT	5640
10	GCTAAAGATG CGACTACAAG TTTAATGAGT CGCTTAACAA CGTTTTTAGA GCATGCACTT	5700
	GCAAAGACGA CATCTTTACA TGGTGTTTTA GTAGATGTTT ACGGTGTTGG TGTACTAATT	5760
	ACCGGTGATT CAGGAATAGG TAAAAGTGAG ACTGCGTTGG AATTAGTTAA ACGTGGGCAT	5820
15	AGATTAGTAG CAGATGATAA TGTAAGAAATA CGTCAAATTA ATAAAGATGA ACTAATAGGG	5880
	AAACCACCAA AGTTAATAGA ACATCTATTA GAAATACGTG GACTAGGTAT TATCAATGTT	5940
20	ATGACTTTAT TTGGCGCGGG TTCAATATTA ACTGAAAAAC GAATTAGATT AAATATTAAT	6000
	TTGGAAAACT GGAACAAGCA AAAGTTATAT GACCGCGTAG GTCTTAATGA AGAGACGCTA	6060
	AGTATTTTAG ATACTGAAAT CACTAAAAAA ACAATACCTG TAAGACCTGG TAGAAATGTT	6120
25	GCGGTAATTA TTGAGTCCG TGCAATGAAC TATCGATTAA ATATCATGGG CATTACACG	6180
	GCCGAAGAAT TTAGTGAAAG ATTAAATGAA GAAATTATCA AGAACAGTCA TAAGAGTGAG	6240
	GAGTAGGTTG AATGGGTATT GTATTTAACT ATATAGATCC TGTGGCATT AACTTAGGAC	6300
30	CACTGAGTGT ACGATGGTAT GGAATTATCA TTGCTGTCGG AATATTACTT GGTACTTTG	6360
	TTgCACAACG TGCACTAGTT AAAGCAGGAT TACATAAAGA TACTTTAGTA GATATTATTT	6420
	TTTATAGTGC ACTATTGGA TTTATCGCGG CACGAATCTA TTTTGTGATT TTCCAATGGC	6480
35	CATATTACGC GGAAAATCCA AGTGAAATTA TTAATAATATG GCATGGTGGA ATAGCAATAC	6540
	ATGGTGGTTT AATAGGTGGC TTTATTGCTG GTGTTATTGT ATGTAAAGTG AAAAAATTAA	6600
	ACCCATTTC AATTGGTGAT ATCGTTGCGC CAAGTATAAT TTTAGCGCAA GGAATTGGAC	6660
40	GCTGGGGTAA CTTTATGAAT CACGAGGCAC ATGGTGGATC GGTGTCACGC GCTTTTTTAG	6720
	AACAATTACA TTTGCCTAAT TTTATAATAG AAAATATGTA TATTAACGGC CAATATTATC	6780
45	ATCCAACATT CTTATATGAA TCCATTTGGG ATGTCGCTGG ATTTATTATC TTAGTTAATA	6840
	TTCGTAAACA TTTAAAATTA GGAGAAACAT TCTTTTTATA TTAACTTGG TATTCAATTG	6900
	GTCGATTCTT TATAGAAGGA TTACGTACAG ATAGCTTAAT GCTCACAAGT AATATTAGAG	6960
50	TTGCACAAIT AGTATCAATT CTTTAAATTT TAATAAGTAT AAGTTTAATT GTATATAGAA	7020
	GGATTAAGTA TAATCCACCG TTSTATAGCA AAGTTGGGGC GCTTCCATGG CCAACAAAAA	7080
55		

	TTATGGCGTG TATACCGTCT TGTAAATTT TCGAAAGTTT TTAAGAATGT AATTATCATT	7200
	GAATTTTCGA AATTTATTCC AAGTATGGTA CTGAAAAGAC ATATATATAA ACAACTTTTA	7260
5	AATATTAATA TCGGTAATCA ATCGTCGATA GCTTATAAAG TAATGTTAGA TATTTTITAC	7320
	CCAGAACTGA TTACGATTGG TAGTAACAGT GTTATTGGTT ACAATGTAAC AATTTTGACG	7380
	CATGAAGCAT TAGTTGATGA ATTTGCTTAT GGACCAGTGA CGATAGGATC TAACACTTTG	7440
10	ATTGGTGCAA ATGCTACCAT TTTACCCGGT ATAACGATTG GTGACAATGT AAAAGTTGCA	7500
	GCTGGTACGG TTGTTTCAAA AGATATACCG GATAATGGAT TTGCATATGG CAACCCTATG	7560
	TATATAAAAA TGATTAGGAG GTGACAATTT TATGGCGCAA AAGAATAATA ATGTAATTCC	7620
15	AATGACTTTT GATGATGCAT TTTATCGTAA AATGGCTAAA CAGAAGTTTA AACAAAGAGA	7680
	ATATAAACGA GCTGCTGAAT ACTTTGAAAA AGTGTTAGAA TTGTCACCTG ATGATCTGGA	7740
20	AATTCAAATT GATTATGCAC AATGTCTAGT GCAACTTGGT ATTGCTAAAA AAGCAGAACA	7800
	TTTATTTTAT GACAAATATTA TTTATAATAG GCATCTAGAA GATAGCTTTT ATGAATTGAG	7860
	TCAGCTCAAC ATTGAAGTTA ACGAACCAAA CAAGGCATTC TTGTTTGCTA TTAATTATGT	7920
25	TATTGTTAGC GACGACCAAG ATTATAGAGA TGAATTAGAT CAAATGTTTG ATGTGAAATA	7980
	TCAAAGTGAA GAACAAATTG AACTTGAAGC TCAATTGTTT GTAGTTCAAA TACTATTCCA	8040
	ATATCTTTTT TCTCAAGGTC GATTAAAAGA TGCAAAGAAT TATGTCTTAC ATCAACCACA	8100
30	AGAAGTTCAA GATCATCGTG TAGTACGTAA TTTATTGGCA ATGTGTTATT TATATCTCGG	8160
	TGAATATGAT ACgGCTAAAG CATTGTACGA aGCACTATTA CAAGAGGATA GTACaGATAT	8220
	ATATGCATTA TGCCATTATA CTTTGCTACT TTATAACACT AAGGAAAATG AACAAATATCA	8280
35	AAAATATTTA AAAATATTAA ACAAAGTTGT ACCTATGAAT GACGATGAAA GTTTTAAATT	8340
	AGGTATTGTA TTAAGTTATT TAAAGCAGTA TCGTGCATCA CAACAATTGT TGTACCCTTT	8400
40	ATATAAAAAA GGGAAATTTT TATCAATTCA AATGTACAAT GCTTTAGCAT ATAATTATTA	8460
	TTATTTAGGT GAAGAAGACG AAAGTCATTA CTA CTGGGAT AAATTGAAGC AAATTTCTAA	8520
	AGTGGAATTT GGACATGCGC CTTGGGTAAT TGAAAATAGC AAAGAAGTTT TTGACCAACA	8580
45	TATTTTGCCA TTACTTCAAA GTGATGACAG TCATTATCGT TTATATGGTA TTTTTTTATT	8640
	GGATCAATTA AATGGTAAAG AAATTGTGAT GACGGAAAGT ATTTGGCAGG TTTTGAAAA	8700
	TCTAAATAAT TATGAGAAAT TGTATTTAAC GTATTTAGTT CAAGGTTTAA CGCTCAATAA	8760
50	ATTAGACTTC ATTCATCGCG GCTTATTAAC GCTTTACCAT AATGAATTAT TTGTAAGTGA	8820
	AAATGATGTA ATGTTTGCAT GGATTAATCA AGGTGAACTC ATAATTGCTG AAAAAAGTAGA	8880
55		

## EP 0 786 519 A2

	TCGAAACGTT	ACAAAGAAGC	AAATTACAAC	ATGGTTAGGC	ATAACACAAT	ATAAACTGAA	9000
	CAAAATGATT	GAATTTCTCT	TGAGCATATA	GATTTATGAA	AAGTTAGATT	TATTATATAA	9060
5	TGCGCATAAT	GATTAATAAT	GAGGAGGCGT	TAATAAAATG	ACTGAAATAG	ATTTTGATAT	9120
	AGCAATTATC	GGTGCAGGTC	CAGCTGGTAT	GACTGCTGCA	GTATACGCAT	CACGTGCTAA	9180
	TTTAAAAACA	GTTATGATTG	AAAGAGGTAT	TCCAGGCGGT	CAAATGGCTA	ATACAGAAGA	9240
10	AGTAGAGAAC	TTCCCTGGTT	TCGAAATGAT	TACAGGTCCA	GATTTATCTA	CAAAAAATGTT	9300
	TGAACACGCT	AAAAAGTTTG	GTGCAGTTTA	TCAATATGGA	GATATTAAAT	CTGTAGAAGA	9360
	TAAAGGCGAA	TATAAAGTGA	TTAACTTTGG	TAATAAAGAA	TTAACAGCGA	AAGCGGTTAT	9420
15	TATTGCTACA	GGTGCAGAAT	ACAAGAAAAT	TGGTGTTCCT	GGTGAACAAG	AACTTGGTGG	9480
	ACGCGGTGTA	AGTTATTGTG	CAGTATGTGA	TGGTGCATTC	TTTAAAAATA	AACGCCTATT	9540
20	CGTTATCGGT	GGTGGTGATT	CAGCAGTAGA	AGAGGGAACA	TTCTTAACTA	AATTTGCTGA	9600
	CAAAGTAACA	ATCGTTCACC	GTCTGTATGA	GTTACGTGCA	CAGCGTATTT	TACAAGATAG	9660
	AGCATTCAAA	AATGATAAAA	TCGACTTTAT	TTGGAGTCAT	ACTTTGAAAT	CAATTAATGA	9720
25	AAAAGACGGC	AAAGTGGGTT	CTGTGACATT	AACGTCTACA	AAAGATGGTT	CAGAAGAAAC	9780
	ACACGAGGCT	GATGGTGTAT	TCATCTATAT	TGGTATGAAA	CCATTAACAG	CGCCATTTAA	9840
	AGACTTAGGT	ATTACAAATG	ATGTTGGTTA	TATTGTAACA	AAAGATGATA	TGACAACATC	9900
30	AGTACCAGGT	ATTTTTCAG	CAGGAGATGT	TCGCGACAAA	GGTTTACGCC	AAATTGTCAC	9960
	TGCTACTGGC	GATGGTAGTA	TTGCAGCGCA	AAGTGCAGCG	GAATATATTG	AACATTTTAA	10020
	CGATCAAGCT	TAATTCGAAG	TCGAATTAAG	ATGTTGAGCT	GTAATTTATT	TGGATATTTA	10080
35	TTTTAATAGT	GTATCACAG	CGTTAAATA	ATGTCCTACT	TTTAAATTAA	AGCAAATTAT	10140
	ATAGAAAACT	AGAACTTAGT	ACGTATCATT	TGTGCGTTTC	AATGAGTTCT	AGTTTTTTTA	10200
	TATGTTATAT	TAACTTATA	ACTTTATGGG	AGTGGGACAG	AAATGATAAA	GAGCCACTAA	10260
40	TGATTTATTA	TGTAGTGGTT	CTTAAACATT	AGCCACAGCT	AATGTGTACT	TAAAAATAGG	10320
	AATACATGAG	TAAAACTCAT	GCATAAGAAA	TACTAATTTT	TATAGAAAAA	GTATTACTTT	10380
45	ATCGTTGTCC	CACCCCAACT	TGCACATTAT	TGTAAGCTGA	CTTCCGCCA	GCTTCTGTGT	10440
	TGGGGCCCCG	CCAACCTGCA	CATTATTGTA	AGCTGACTTT	TCGTCAgCTT	CTGTGTTGGG	10500
	GCCCCGCCAA	CTTGACACATT	ATTGTAAGCT	GACTTTTCGT	CAGCTTCTGT	GTTGGGGCCC	10560
50	CGCCAACTTG	CATTGTCTGT	AGAAATTGGG	AATCCAATTT	CTCTATGTTG	GGGCCCACAC	10620
	CCCAACTCGC	ATTGCCTGTA	GAATTTCTTT	TCGAAATTCT	CTGTGTTGGG	GCCCACACCC	10680
55							

	ACTCGCATTG CCGTAGAAT TTCTTTTCGA AATTCTCTGT GTTGGGGCCC CTGACTAGAG	10800
	TTGAAAAAAG CTTGTTGCAA GCGCAITTTT ATTCACTCAA CTACTAGCAA TATAATATTA	10860
5	TAGACCCTAG GACATTGATT TATGTCCCAA GCTCCTTTTA AATGATGTAT ATTTTGTAGAA	10920
	ATTTAATCTA GACATAGTTG GAAATAAATA TAAACATCG TTGCTTAATT TTGTCATAGA	10980
	ACATTTAAAT TAACATCATG AAATTCGTTT TGGCGGTGAA AAAATAATGG ATAATAATGA	11040
10	AAAAGAAAA AGTAAAAGTG AACTATTAGT TGTAACAGGT TTATCTGGCG CAGGTAAATC	11100
	TTTGGTTATT CAATGTTTAG AAGACATGGG ATATTTTGT GTAGATAATC TACCACCAGT	11160
	GTTATTGCCT AAATTTGTAG AGTTGATGGA ACAAGGAAAT CCATCCTTAA GAAAAGTGGC	11220
15	AATTGCAATT GATTTAAGAG GTAAGGAACT ATTTAATTCA TTAGTTGCAG TAGTGGATAA	11280
	AGTCAAAAGT GAAAGTGACG TCATCATTGA TGTATGTTT TTAGAAGCAA GTACTGAAAA	11340
20	ATTAATTTCA AGATATAAGG AAACGCGTCG TGCACATCCT TTGATGGAAC AAGGTAAAAG	11400
	ATCGTTAATC AATGCAATTA ATGATGAGCG AGAGCATTTG TCTCAAATTA GAAGTATAGC	11460
	TAATTTTGTT ATAGATACTA CAAAGTTATC ACCTAAAGAA TTAAGAAGAAC GCATTCGTCG	11520
25	ATACTATGAA GATGAAGAGT TTGAACTTT TACAATTAAT GTCACAAGTT TCGGTTTAA	11580
	ACATGGGATT CAGATGGATG CAGATTTAGT ATTTGATGTA CGATTTTAC CAAATCCATA	11640
	TTATGTAGTA GATTTAAGAC CTTTAACAGG ATTAGATAAA GACGTTTATA ATTATGTTAT	11700
30	GAAATGGAAG GAGACGGAGA TTTTCTTGA AAAATTAATC GATTTGTTAG ATTTTATGAT	11760
	ACCCGGGTAT AAAAAAGAAG GGAAATCTCA ATTAGTAATT GCCATCGGTT GTACGGGTGG	11820
	ACAACATCGA TCTGTAGCAT TAGCAGAACG ACTAGGTAAT TATCTAAATG AAGTATTTGA	11880
35	ATATAATGTT TATGTGCATC ATAGGGACGC ACATATTGAA AGTGGCGAGA AAAAATGAGA	11940
	CAAAFAAAAG TTGTACTTAT CGGTGGTGGC ACTGGCTTAT CAGTTATGGC TAGGGGATTA	12000
	AGAGAATTCC CAATTGATAT TACGGCGATT GTAACAGTTG CTGATAATGG TGGGAGTACA	12060
40	GGGAAAATCa GAGATGAAAT GGATATACCA GCACCAGGAG ACATCAGAAA TGTGATTGCA	12120
	GCTTTAAGTG ATTCTGAGTC AGTTTAAAGC CAACTTTTTC AGTATCGCTT TGAAGAAAAT	12180
45	CAAATTAGCG GTCATCATT AGGTAATTTA TTAATCGCAG GTATGACTAA TATTACGAAT	12240
	GATTTGCGAC ATGCCATTAA AGCATTAAGT AAAATTTTAA ATATTAAAGG TAGAGTCATT	12300
	CCATCTACAA ATACAAGTGT GCAATTAAAT GCTGTTATGG AAGATGGAGA AATTGTTTTT	12360
50	GGAGAAACAA ATATTCCTAA AAAACATAAA AAAATTGATC GTGTGTTTTT AGAACCTAAC	12420
	GATGTGCAAC CAATGGAAGA AGCAATCGAT GCTTTAAGGG AAGCAGATTT AATCGTTCTT	12480

55

GCGTTAATTC ATTCTGATGC GCCTAAGCTA TATGTTTCTA ATGTGATGAC GCAACCTGGG 12600  
 GAAACAGATG GTTATAGCGT GAAAGATyAT ATCGATGCCA TTCATAGACA AGCTGGACAA 12660  
 5 CCGTTTATTG ATTATGTCAT TTGTAGTACA CAAACTTTCA ATGCTCAAGT TTTGAAAAAA 12720  
 TATGAAGAAA AACATTCTAA ACCAGTTGAA GTTAATAAGG CTGAACTTGA AAAAGAAAGC 12780  
 ATAAATGTAA AAACATCTTC AAATTTAGTT GAAATTTCTG AAAATCATT T AGTAAGACAT 12840  
 10 AATACTAAAG TGTTATCGAC AATGATTAT GACATAGCTT TAGAATTAAT TAGTACTATT 12900  
 CCTTTCGTAC CAAGTGATAA ACGTnAATAA TATAGAACGT AATCATATTA TGATATGATA 12960  
 ATAGAGCTGT GAAAAAATG AAnATAGACA GTGGTTCTAA GGTGAATCAT GTTTTAAATA 13020  
 15 AGAAAGGAAT GACTGTACGA TGAGCTTTCG ATCAGAAATG AAAAATGAAT TAACTAGAAT 13080  
 AGACGT 13086

## (2) INFORMATION FOR SEQ ID NO: 17:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1350 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 25 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

30 CATTAGTCAT GAAATAGCC GACAACTTCA TCTGTGAAAT CACCGGCCTT TTATTTTAGC 60  
 TAACTTTATT TCTGATTTTA CGATTTTAAT TGATCATACA GAGAAAGTGA TCTTTTACAA 120  
 ATTTCTAAAA ACTCATGATC TATATTGGAC ATTTGATGAA AATAAGACAA AATGTTTCT 180  
 35 GTTAGCTTCT CTTGTTTTGG GAATGAATCA TCCTCTTTAA TCCAAATCGC TAATTCGCCT 240  
 AATGTTGTTT TATCATCTTT AAATGTTTGT ATATATTCGT AAAAGCTCAT AGTATTCCTT 300  
 CTCTCAATTT ACTTATATAA ATCCTACCAC GAAAGCTTTC AAGAAAACAC AATTAAATGT 360  
 40 CTATTTAGTG AACTTTTAA GGTGTGCAC TCTTTAATG TCTGCCAATT AGGTCAATTA 420  
 ATCATCACAA TGTACAATTA ACTCTATTTT CAGTTCATAT ACTCACACAC CGTTTTTGAA 480  
 45 CAACACATTA ACTTCTCATT TAGATAAAAC GCAAAAAAGC CTGGCACCAA TACAATAGAT 540  
 GCCAGACTAA GAGTCTACTA TATAAATTTA TTTAGCGTAT GGTTTTACTT CGATTGCACC 600  
 TTCATTTTCA TCATGAACAC CATGCTTATA ATAATCAATA TATTGTGGCT CTAAAGGCTT 660  
 50 TCTGCCACGT ATAATGTCTG CTGCTTTTTC AGCTAACATT AAAACAGGTG CGTGTATATT 720  
 GCCATTTGTC GTACGTGGCA TAGCTGATGC ATCAACTACA CGTAAATTTT CCATACCGTG 780

ACTACAAGAT GGGTGAATG CTGTTTCACC ATCTCTACGA ACCCAATCAA GAATTTCTTC 900  
 GTCTGTTTGC ACTTCTGGTC CTGGTGAAAT TTCTCCACCA TTGAATGGAT CCATTGCTTT 960  
 5 TTGAGATAAG ATATTTCTTG CTACACGAAT TGCTTCTACC CATTCTTTTT TATCTTCTTC 1020  
 TGTGATAAAA TAATTAAAGC GGATACTTGG TTTTTCGAAT GGATCTTTAG ATTTGATTTT 1080  
 CAAGCTACCA CGAGAGTTTG AATACATTGG TCCTACGTGA ACTTGATAAC CATGTGCGAC 1140  
 10 CGCTGCCTTT TGACCATCAT ATCTTACAGC TATTGGTAAG AAATGGAACA TTAAGTTAGG 1200  
 ATAAtCAACT TCGTTATTTG AACGTACAAA TCCGCCACCT TCAAAATGGT TAGATGCTGC 1260  
 15 TGCACCTGTA CGTGTGAAAA TCCATTGTAA ACCAATAAAT GGcATGCGCT TGAtATCTAA 1320  
 GCTTGGCtGt AATGATACAG GTTCCTTACA 1350

## (2) INFORMATION FOR SEQ ID NO: 18:

20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1376 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

TAATGCTATT GGCAACACCA TATATGAAAn CTCCAAACGA TCCTAAACCG ACTATAGATT 60  
 30 CACCAAATTT nACAATCCAT GAATAAAGTA GTGGCCATAA GAATAACAAT ATGACAACCTA 120  
 AAAATGTACA GTAAAATGCA GTCATAATTG GAACTAGACG TTTACCACTA AAAAATGATA 180  
 ATGCTAATGG TAATTCTGTT TCACTAAACT TATTGTATGC ATAAGCTGCT ATTAAACCTA 240  
 35 TTACAATACC AACAAAGACA TTGCCATTAT TCATCTTTTC AAAAGCTGAA TTTATTTCCG 300  
 ArGCTTTTCAT TCCTAATAAA GGCGCTAATT TCATTGGTGA TAATACAACCT GTAACATAAA 360  
 40 AATATCCTAA CGTrGCTGCA rGCGsGACTG CACCATCATT TTTCTTTGCC ATTCCTATAG 420  
 CTACACCAAT TGCAAATAAA ATACCTAATT GCTCTAAAAT CGTAGTACCT ACCGTAGTAA 480  
 AGAACATTGC GATTTTCGGC GTCGCATGAA GTGCATTTAA CGTATTACCA ATTCCGGCAA 540  
 45 TAATTGCTGC AGCCGGTAAA ATGGCAACTG GTAACATTAA CGAACGCCCT AAATTTTGGA 600  
 AAAATTTATA CATTGAATGT CATCCTTCTT AAAATAATGT AGAAATATAA AGATTACTAA 660  
 TGTAAC TAGA ATAAC TACTT CGATACTCCG TTATAGTCAC CTAGGCTTAC TAACCAGCTA 720  
 50 TATTTCTACC TCAAGTTATT TTATAAACTT TTTACAATTT CATGCAATTC TTGTTGTAAC 780  
 TTTGCTGTTT GTGTTTCAAT CTCTTTTGTA ATATAATCGA TACGCTCGTT TCGTTTTAAA 840

55



AAAGACCGTG AATCTTAGTA GGACCAACAT AAGCAACAGG TAATATTGGT GACTTACTTA 960  
 ACATTGCAAT TGTTGAAGCA CCaCGTTTCA AAGGTGCACC TTCTTGCGAT GTGCGAGAAC 1020  
 5 CTGTTGGGAA GATACCAACT GTCTTATTAT CTTTCAACAA ATTGATTGGG CGTTTTAAAG 1080  
 TACTAGGTCC TGGATTTTCA CGATCTACAG GAAATGCATT TAAAGACGTT AAAAATTTAC 1140  
 CAATCCATTT ATTTTTGAAT AATTCTTTTT TAGCCATATA ATGAATTGGA TTAGGATATA 1200  
 10 ATGCCATACC TAGCATAATG ACTTCGTTAT AACTTTCATG CGTACAAGTT ACGACATATT 1260  
 TACTATCCTT AGGAATATTA TCTTTACCGA TTACGTATAA TGATTTTGAC ATTTTAACTA 1320  
 AAATGAAATT CAAAATCTTA CTAATCACTG AATACATTGT GCCACCTACT TAACTT 1376

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7363 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

TTGTCATACC AATATTTTGT AAAATATGGA ACACAAGTAA AGTGACGAAA CCAACGATAA 60  
 AGATTTTGTT AAATTGATCT TCAATTTTCG CAGCTAATCT TATTAGATGG AAGATTAAAA 120  
 30 ATAAAAATAT TAAGATCAAT ATGACAGAAC CGATAAAGCC AAGTTCCTCT CCAATCACTG 180  
 AAAAGATAAA GTCAGTATGA TTTTCAGGTA TATAAACTTC ACCGTGATTG TATCCTTTAC 240  
 CTAGTAAC TG CAGAACCG ATAGCTTTAA GTGATTCAGT TAAATGaTAG CCATCACCAC 300  
 35 TACTATATGT ATAGGGGTCA AGCCATGAAT TGATTGCTCC CATTTGATAC AGTTGGaCAC 360  
 CTAATJAAATT TTCAATTAAT GCGGGTGCAT ATAGaATACC TAAAATGACT GTCATTGCAC 420  
 CAACaATACC TGTAATAAAG ATAGGTGCTA AGATACGCCA TGTTATACCA CTTACTAACA 480  
 40 TCACACCTGC AATAATAGCA GCTAATACTA ATGTAGTTCC TAGGTCATTT TGCAGTAATA 540  
 TTAATAACT TGGTACTAAC GAGACACCAA TAATTTTGAA AAATAATAAC AAATCACTTT 600  
 GGAATGATTT ATTGAATGTG AATTGATTAT GTCTAGAAAC GACACGCGCT AATGCTAAAA 660  
 45 TTAATAAAT TTTGATGAAT TCAGATGGCT GAATACTGAT AGGGCCAAAC GTGTACCAAC 720  
 TTTTGGCACC ATTGATAATA GGTGTAATAG GTGACTCAGG AATAACGAGC AAGCCTATTA 780  
 50 ATAATAGACA GATTAAGAAA TACAATAAAT ATGTATAATG TTTAATCTTT TTAGGTGAAA 840  
 TAAACATGAT GATACCTGCA AAAATTGCAC CTAAAATGTA ATAAAAAATT TGTCTGATAC 900

	TTGCTAAAAC AGCTATAGTG GCTACTAATA CCCAGTCTAC TTTGCGAAAC aATGCTTATC	1020
	CGGCTGTTGA CGAGATGAAT AATTCATTGC AAACCTCCTT TATACTCACT AATGTTTATA	1080
5	TCAATTTTAC ATGACTTTTT AAAAATTAGC TAGAATATCA CAGTGATATC AGCTATAGAT	1140
	TTCAATTTGA ATTAGGAATA AAATAGAAGG GAATATTGTT CTGATTATAA ATGAATCAAC	1200
	ATAGATACAG ACACATAAGT CCTCGTTTTT AAAATGCAAA ATAGCATTAA AATGTGATAC	1260
10	TATTAAGATT CAAAGATGCG AATAAATCAA TTAACAATAG GACyAAATCA ATATTAATTT	1320
	ATATTAAGGT AGCAAACCTT GATATATCAT TGGAGGAAAA CGAAATGACA AAAGAAAATA	1380
	TTTGTATCGT TTTTGGAGGG AAAAGTGCAG AACACGAAGT ATCGATTCTG ACAGCACAAA	1440
15	ATGTATTAAA TGCAATAGAT AAAGACAAAT ATCATGTTGA TATCATTAT ATTACCAATG	1500
	ATGGTGATTG GAGAAAGCAA AATAATATTA CAGCTGAAAT TAAATCTACT GATGAGCTTC	1560
20	ATTTAGAAAA TGGAGAGGCG CTTGAGATTT CACAGCTATT GAAAGAAAGT AGTTCAGGAC	1620
	AACCATACGA TGCAGTATTC CCATTATTAC ATGGTCCTAA TGGTGAAGAT GGCACGATTC	1680
	AAGGGCTTTT TGAAGTTTTG GATGTACCAT ATGTAGGAAA TGGTGTATTG TCAGCTGCAA	1740
25	GTTCTATGGA CAAACTTGTA ATGAAACAAT TATTTGAACA TCGAGGGTTA CCACAGTTAC	1800
	CTTATATTAG TTTCTTACGT TCTGAATATG AAAAATATGA ACATAACATT TTAAAATTAG	1860
	TAAATGATAA ATTAAATTAC CCAGTCTTTG TTAAACCTGC TAACTTAGGG TCAAGTGTAG	1920
30	GTATCAGTAA ATGTAATAAT GAAGCGGAAC TTAAAGAAGG TATTAAAGAA GCATTCCAAT	1980
	TTGACCGTAA GCTTGTATA GAACAAGGCG TTAACGCACG TGAAATTGAA GTAGCAGTTT	2040
	TAGGAAATGA CTATCCTGAA GCGACATGGC CAGGTGAAGT CGTAAAAGAT GTCGCGTTTT	2100
35	ACGATTACAA ATCAAAATAT AAAGATGGTA AGGTTCAATT ACAAATTCCA GCTGACTTAG	2160
	ACGAAGATGT TCAATTAACG CTTAGAAATA TGGCATTAGA GGCATTCAAA GCGACAGATT	2220
	GTTCTGGTTT AGTCCGTGCT GATTTCTTTG TAACAGAAGA CAACCAAATA TATATTAATG	2280
40	AAACAAATGC AATGCCTGGA TTTACGGCTT TCAGTATGTA TCCAAAGTTA TGGGAAAATA	2340
	TGGGCTTATC TTATCCAGAA TTGATTACAA AACTTATCGA GCTTGCTAAA GAACGTCACC	2400
45	AGGATAAACA GAAAAATAAA TACAAAATTG ACTAACTGAG GTTGTATTATA TGATTAATGT	2460
	TACATTAAAG CAAATTCAAT CATGGATTCC TTGTGAAATT GAAGATCAAT TTTTAAATCA	2520
	AGAGATAAAT GGAGTCACAA TTGATTCACG AGCAATTCTT AAAAATATGT TATTTATACC	2580
50	ATTTAAAGGT GAAATGTTG ACGGTCATCG CTTTGTCTCT AAAGCATTAC AAGATGGTGC	2640
	TGGGGCTGCT TTTTATCAAA GAGGGACACC TATAGATGAA AATGTAAGCG GGCCTATTAT	2700

55

	AAACCCTAAA	GTAATTGCCG	TCACAGGGTC	TAATGGTAAA	ACAACGACTA	AAGATATGAT	2820
	TGAAAGTGTA	TTGCATACCG	AATTTAAAGT	TAAGAAAACG	CAAGGTAATT	ACAATAATGA	2880
5	AATTGGTTTA	CCTTTAACTA	TTTTGGAATT	AGATAATGAT	ACTGAAATAT	CAATATTGGA	2940
	GATGGGGATG	TCAGGTTTCC	ATGAAATTGA	ATTTCGTGCA	AACCTCGCTC	AACCAGATAT	3000
	TGCAGTTATA	ACTAATATTG	GTGAGTCACA	TATGCAAGAT	TTAGGTTTCGC	GCGAGGGGAT	3060
10	TGCTAAAGCT	AAATCTGAAA	TTACAATAGG	TCTAAAAGAT	AATGGTACGT	TTATATATGA	3120
	TGGCGATGAA	CCATTATTGA	AACCACATGT	TAAAGAAGTT	GAAAATGCAA	AATGTATTAG	3180
	TATTGGTGTT	GCTACTGATA	ATGCATTAGT	TTGTTCTGTT	GATGATAGAG	ATACTACAGG	3240
15	TATTTCAATTT	ACGATTAATA	ATAAAGAACA	TTACGATCTG	CCAATATTAG	GAAAGCATAA	3300
	TATGAAAAAT	GCGACGATTG	CCATTGCGGT	TGGTCATGAA	TTAGGTTTGA	CATATAACAC	3360
20	AATCTATCAA	AATTTAAAAA	ATGTCAGCTT	AACTGGTATG	CGTATGGAAC	AACATACATT	3420
	AGAAAATGAT	ATTACTGTGA	TAAATGATGC	CTATAATGCA	AGTCCTACAA	GTATGAGAGC	3480
	AGCTATTGAT	ACACTGAGTA	CTTTGACAGG	GCGTCGCATT	CTAATTTTAG	GAGATGTTTT	3540
25	AGAATTAGGT	GAAAATAGCA	AAGAAATGCA	TATCGGTGTA	GGTAATTATT	TAGAAGAAAA	3600
	GCATATAGAT	GTGTTGTATA	CGTTTGGTAA	TGAAGCGAAG	TATATTTATG	ATTCGGGCCA	3660
	GCAACATGTC	GAAAAAGCAC	AACACTTCAA	TTCTAAAGAC	GATATGATAG	AAGTTTTAAT	3720
30	AAACGATTTA	AAAGCGCATG	ACCGTGTATT	AGTTAAAGGA	TCACGTGGTA	TGAAATTAGA	3780
	AGAAGTGGTA	AATGCTTTAA	TTTCATAGAG	ATTAGTCGAG	GGACCTTTTA	CTTATAAAAA	3840
	TGATTTGAAT	TAATACTAAA	AGATTACAAA	GAAGAGGTGG	TTTTGTGTGT	AAATACAAAA	3900
35	TTGCCTTTTT	CTTTTTATGT	TAAATCTATA	AATTTGAAAC	TAAATCAAGG	TTAATTCTAT	3960
	GTACACACTT	TATATAGGAA	GTAGTTTGAA	TGTTTATATA	ATGTTTTTACA	AAAAGATGTA	4020
	GTATTATAAT	GTCTAATTTT	ACATGTGTTT	CAGTAAATTT	TGTTGTGGAA	TGTTAACCAT	4080
40	ATACGTATTT	TATAAAAAAT	TTTTTATAAT	GATTATTCTG	ATGATGCGTA	ACGCTTACAT	4140
	CTTATCTAAT	GCTAGCTTTT	TGACAAAAAT	ATGACAATCA	ATTAATGTGA	TTCTAATAAA	4200
45	TATTCGCAAA	TTGCTTTATT	GCGATTAAAT	TTTTTTGGTG	GTAATATATA	GAAGTTGATG	4260
	AAATATTAAT	GAACCTTATAT	GCAAAAAGTAT	ATTGAGAAAT	AAACAGGTAA	AAAGGAGAAT	4320
	TATTTTGCAA	AATTTTAAAG	AACTAGGGAT	TTCCGATAAT	ACGGTTCAGT	CACTTGAATC	4380
50	AATGGGATTT	AAAGAGCCGA	CACCTATCCA	AAAAGACAGT	ATCCCTTATG	CGTTACAAGG	4440
	AATTGATATC	CTTGGGCAAG	CTCAAACCGG	TACAGGTAAA	ACAGGAGCAT	TCGGTATTCC	4500

	AGAATTGGCA ATGCAGGTAG CTGAACAATT AAGAGAATTT AGCCGTGGAC AAGGTGTCCA	4620
	AGTTGTTACT GTATTCGGTG GTATGCCTAT CGAACGCCAA ATTAAAGCCT TGAAAAAAGG	4680
5	CCCACAAATC GTAGTCGGAA CACCTGGGCG TGTTATCGAC CATTTAAATC GTCGCACATT	4740
	AAAAACGGAC GGAATTCATA CTTTGATTTT AGATGAAGCT GATGAAATGA TGAATATGGG	4800
	ATTCATCGAT GATATGAGAT TTATTATGGA TAAAATTCCA GCAGTACAAC GTCAACAAT	4860
10	GTTGTTCTCA GCTACAATGC CTAAAGCAAT CCAAGCTTTA GTACAACAAT TTATGAAATC	4920
	ACCAAAAATC ATTAAGACAA TGAATAATGA AATGTCTGAT CCACAAATCG AAGAATTCTA	4980
	TACAATTGTT AAAGAATTAG AGAAATTTGA TACATTTACA AATTTCCCTAG ATGTTTCATCA	5040
15	ACCTGAATTA GCAATCGTAT TCGGACGTAC AAAACGTCGT GTTGATGAAT TAACAAGTGC	5100
	TTTGATTCTT AAAGGATATA AAGCTGAAGG TTTACATGGT GATATTACAC AAGCGAAACg	5160
20	TTtAGAAGTA TTanAGAAAT TTA AAAATGA CCAAAATTAAT ATTTTAGTCG CTA CTGATGT	5220
	AGCAGCaAGA GGA CTAGATA TTTCTGGTGT GAGTCATGTT TATAACTTTG ATATACCTCA	5280
	AGATACTGAA AGCTATACAC ACCGTATTGG TCGTACGGT CGTGCTGGTA AAGAAGGTAT	5340
25	CGCTGTAACG TTTGTTAATC CAATCGAAAT GGATTATATC AGACAAATTG AAGATGCAAA	5400
	CGGTAGAAAA ATGAGTGCAy TcGTCCACCA CATCGTAAAG AAGTACTTCA AGCAGTGAA	5460
	GATGACATCA AAGAAAAAGT TGAAAAC TGG ATGTCTAAAG AGTCAGAATC ACGCTTGAAA	5520
30	CGCATTTCTA CAGAGTTGTT AAATGAATAT AACGATGTTG ATTTAGTTGC TGCACTTTTA	5580
	CAAGAGTTAG TAGAAGCAAA CGATGAAGTT GAAGTTCAAT TAAC TTTTGA AAAACCATTA	5640
	TCTCGCAAAG GCCGTAACGG TAAACCAAGT GGTTCTCGTA ACAGAAATAG TAAGCGTGGT	5700
35	AATCCTAAAT TTGACAGTAA GAGTAAACGT TCAAAAGGAT ACTCAAGTAA GAAGAAAAGT	5760
	ACA AAAAAAT TCGACCGTAA AGAGAAGAGC AGCGGTGGAA GCAGACCTAT GAAAGGTCGC	5820
	ACATTTGCTG ACCATCAAAA ATAATTTATA GATTAAGAGC TTAAAGATGT AATGTCTTGA	5880
40	GCTCTTTTTT GTTTTCAATA ATTGATTCTC TG TAGATATC aAAGTaCTAA CGTTTTAAAG	5940
	GT TAAATATT TAATTGGATT GAGATCTGTA TGCGGT TATA TCaTTCTGTG TAAATATGGT	6000
45	TCTCCACCAA ATGTGGTGAG TATATAATTT AAAGAACTAT TTTTAAATTA AGAATAATCG	6060
	AACATAAATA AACTTTATGA AATTT CAGTA TCATGTTCTT ATAAAAACA ATAGGGCTTT	6120
	TTGctGACGC TAGTGCGGA TAAATAATAA GTTGAATATA AAAAAGATCA CTGCCAATCA	6180
50	TTCGTTTAAT GGCAGCGATC TTTTTTATTT AATTATTCT CTTTCCACTG CAACATTTGA	6240
	TAACCAATGC GTGGATGTGT TTAAATAATA TCTTTTGCGT CCTCATGACA TTGTGAAAGT	6300
55		

CCATATATTC GTTTTAATAT CATCTCATAA GTGACTACTT TTCCTTTATG ATTTGACAAT 6420  
 AGTTCTAACA AGCTAAATTC ATTTGGCGTC AAATGTACCT CCTGATTATT AATAACAACA 6480  
 5 GATTTGGAGC CAAAGTCGAT GCTTAGCAAA CCGTTAGTAA ATACAATGTT AGTTTCTTGA 6540  
 TGTGACTTAG CGATTCTCTC GATGACTCGT ATTGCTGCCC GAAGCTCATC AACATTAAAA 6600  
 10 GGTITAGTCA TATAGTCATT CGCACCCTTA TCTAAAGCTT GAATAATTGT TTGTTCTTCT 6660  
 TGTCTTGCAC TTATTACAAT GATAGGAATG TCAGTATGTT GCCTGATTTC TGAAATCAAA 6720  
 CATAATCCAT CTTTATCTGG TAAACCTAAA TCTAATAAAA TGACATCTGG TTTATCAATT 6780  
 15 TGAATTTTAA AGTGTGCTTG TGTGGCATTG TCGGCTGTAG TTACATTGTA ATAATCTAAA 6840  
 GTTAATGCAA CATCAAGTAA ATGTGTGATT GCGTGATCAT CTTCAATTAT CAATATTTTA 6900  
 GATTGCATTA TACGTCTCCT TCGTTAAAGT CTGTATATAT ATTGAAATAG AATATACTGC 6960  
 20 CGTGTGGTTG GTTCGGTTTA TATTGTAAGT TTGATTGATG TTTGTGTAGG ATAGTCTGTA 7020  
 CTAAATATAA GCCTAGTCCC ATGCTTTCTT TTTGGTTATC TTTAAAATAT TTATTTGATC 7080  
 CTGTGTAAAA AGGCTCGAAT ATCTTTTGTt GTTCTTCTAA ACTAATTCCA GGTCTTCTGT 7140  
 25 CTATAACGGC AAATTCGATT TGTTCATAGC TAGCATAACG AATAGATAAA TTGATTTTGG 7200  
 TGTCAGTAGA AGTGTGTTTA ACTGCATTTT CAATCAAATT GAAtAAAgCT TGTAAATCA 7260  
 ACTTACTGTC AATGTGTATA AAcGTAAAT TTAGTGAGGA TGATACAGTT ATACGCTTTT 7320  
 30 TTAAATGGCG ACGTTCTAAA ATACATATCG ATTTCTTATA CTA 7363

## (2) INFORMATION FOR SEQ ID NO: 20:

## (i) SEQUENCE CHARACTERISTICS:

- 35 (A) LENGTH: 10470 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

TTAACAATCG ATAACCACAA TACTTCTATT GTAATTGTTT AACGATTTCn CGATTAAAAAT 60  
 45 CATCTAAATC GTCTGGTACT CGACTTGTTA CAATATTGTT GTCTACAcTa CTGACTCATC 120  
 AACTACATGT GCGCCTGCAT TTGATAAATC TTTGCGTACA TTAAATACTG CTGTTAACGT 180  
 ACGACCTTTT AAATCGTCTG TATCTATTAG TATTTGTGGC CCATGACAAA TGGCAAATGT 240  
 50 TGGTACATCA TTTTtagTAA AGTATTTAGC AAATGTGCCA TATCGACCTT CTGTATCTCC 300  
 ACGTAAATGA TCTGGTGAAA ATCCTCCAGG AATTAATAAT GCATCATAAT CTTCTGGTTT 360

	ATTGTCAGTA TCTCCAATCA CTACAGTATT AAAGCCTGCA TTCTCTAATG CCTCTTTAGG	480
	GCTTGAATAT TCTATATCTT CAAATTGCTT TGCTAGAATA ATTGCTACTT TTTTAGTCAT	540
5	TGAAAATCAC CTTTCTATAT ATCATTGATA TAATTACTAT AGACAAGTAA ATCAGTGATT	600
	AAACATACAA GATATAAAAA ATATTAAGCG ACTGTCGCGA TATCTAACCC TAACACATCT	660
10	TATGTGGCAT TTAATTAGAT ACTAATTAA CCTTTTCTTC AAGCTGATCT AACAAATCCAA	720
	TCCATTCAATC TATATCTTCA ACACGTACTT CATCAGGATT TACATGATCG ATATCCTCAA	780
	TAAACTTATT TAAACGCGCT TTTATCTGTT CGATTGTTTG CTGTTCAATC ATAAAAAGTT	840
15	AACTCCTTTT ATTTTGTTTT CTTTTTCATT ATTATCCTAA CAGAAATTGC GTTAAAGCGA	900
	TATAATCTTA GCTATATTTA TGACATTCAA ATTATTTTGA CTTTAAAAA TCCCCTTTTC	960
	AATTAACTAA AATTAAGAGA TAATTTGTTA CGAGTGATAA TACGAAGkGG TaTCATACCG	1020
20	ATATGAACCA AATAGAAAGA AGGAAGTTTA AGACGATGAA TAGCGTCAAA TTGAAGCAAC	1080
	CTGTTAGCAT TTACAATGAT CCATGGGAAG TGAAATTAT ATACATTTAA ATTTTCATGAG	1140
	ACAATAAAGC TTGATTTAAT GCGTTTTTTT GCCTTTTTTA TTTTCCTTAT TTTTCTGTT	1200
25	TTACAACAAA ATGGTATCAA AAATGGTATC ATTTGTAGTT ATTTTAGCTT CACATATTAA	1260
	AACAACCACA CTCCTAAATT AATAGGTGGT GTGGTTTTGT TGGTTGTGTG GGGATAAAAA	1320
	TAACCGCATC AGTTAAGATG CGGTTATCTA GCAAGGGCCA CGTATTTATA AATACGTTTA	1380
30	GAATCTCTTC GGCAACTTTG CTATAGACAG TCTATGCTGT TACTAAATTA TACCACCACA	1440
	CAAACCTACT CCCATTGAGG AACACAGAGC TTTGTCGCTC GTCAGCAACG TCATATGAAT	1500
	TCTCAGTTCA TGTGTGGTG AACTTTTAAA CGGTCTGTGC CAGTAGCGAC CGAGTCATTT	1560
35	CAAGAATGAC CATTTACAT TTATATTATA ACATTGTGCG TCGGTAACG TATAGTTTTT	1620
	CAGTTGTATT TAAAGTTAAG TTATCTACTT CGCGCTTCC TTGCCTTAAT TGTGAAATTA	1680
	CATATTGCGC TACGCCAGTT TGTTTGTGAA TTTGGTAACC TGTATATCA CTTTGTATCA	1740
40	ATTCAATTAT TTTTAATTTA TAATCACTCA TATTATCTAC GTCCATTCTT TTTATCTAAA	1800
	CAATAAAAAAT GTGTCTTTCT CCCGATAAAT AATAACAATG GTAGGCTTAA TAAAAACAAT	1860
45	ATTAAATACA TTTGTCTGT CATAATTGAA AACCTCCAAA TAATATTATA TTATATAAGT	1920
	GTAAGGAGGA GCCATCAGGC TCCAAGCATA ATGTTAATCT TTGTGTTTG GCTTTCGGTC	1980
	TAGGTAGCCG AGATGCCaTT CTCTAAGTTG TTTTAACACT TCTGGAATTA TCAGTACTGC	2040
50	CAATACTTGA TGTCTAGAA GTGTTTTTAT TATGTCTAGC ATGAGGCTTT TCACCTCCTT	2100
	ACACATAATT TGTAAGTCAT CAACTAACCT ACAAATATAA TTATACTAAA CAAATGTTTA	2160
55		

	GTTATCTACA TTTAAATCTT GAGAGAAATG TTAAAAAGTT CTAATAAAAT AATAGCACAT	2280
	TTTATCTTTA AATGTAAATA GAAAGCAGGT ATGTAACGCA CCTGCTTAAA TAGaCATGAC	2340
5	TATGTCAATC TAACCGATTT CTCCCCATAA GTCACCTAAT ATCTGATTAG GTGGGGCAGA	2400
	ACCATTCAT GTTCTAATAG GCAAGTAATA ACGTTGCCCC TCCCATGTAT ATCCTACCCA	2460
10	AACATGACCA TCTTGTAACA TCACTTCTGT ATAATCACAA TACCCACCAG GTTGGAACGT	2520
	ATAACCCACT GGACAAGATA AGAATGGCCC CACTTTTCTT ACTGTGATTG GTTGATTGCC	2580
	GTTTGTGAAT CTAGCACTTT CTTCATGTA GTAAGTACCA TATTTATTAC GTTTCCATGC	2640
15	ACTTGCAACT GGTTTAACTG TATTACTTGA AGCGCTTGAC TCATTAGAGA CAGTGGCAAC	2700
	CGGTATTTTA CCATCCATGT ACGCCCTAAT CTGCTTGATA AAGTAGTCTT TAAGTTGCAA	2760
	CCGCTTGTCT TCTGGCAATA GACCGCGAGT TACTGGGTCA AAACCAGTGT GTAAAACCGA	2820
20	ACTTCTATGA GGGCATGATG TTGAAGTAAA TTCATGTGC AATCTGATTG TATTTCTGTT	2880
	TGCTGGTAAT CCCCATTTTT TCAACAATCT AGCGCATTCT TGGAAAGTTG CCTGTTCAAT	2940
	TTTTAAGAAT GTCGCGTTAT CTGCGCCCAT TGATTGACAT ACTTCAATAC CGTAATAATA	3000
25	TTTATTACCT ATTGATTAG CGGTATGCCA ACCTACTTGT GATTCATCTA AGGCTTGCCA	3060
	AACTGTGTTG CCGTATACGT AACTATGCGC AATGCCCCGT TCTAATCTTG ATAAAGGTGC	3120
	ATTTACTAAT CCGTTACGAT ATGCTTCAGC AGTCGCCCCT TTGCTCCCTG CGTCGTTGTG	3180
30	TATAACTATA CCTTTAGGGT TACTACCACG CTTAGGTAGG TCATAACCTT TAACCACATC	3240
	TTTGATGATT TTAAGTTCTA CTGCTTTAGG TTGTGGCTTA GCTGTTTCTT TTTTAGGTGC	3300
	TTGTGTAGGA GATTGAACTG ATCGTGGCGC TGTCTCACTT TTAAAATTCTG GACGGATAAA	3360
35	CCACATAGGG AAATCATAAG CATGTTGTCG TCTTGTAAC TTTTCCCAAC CCCAGCCGGG	3420
	TTGTTGCGATT CCGTCAGTCC AGCCACCGCC TAGCCAATTC TGCTCATATA CAATGATGTA	3480
40	ATCTAAAGTT GCTTCAATTA CCCATGCAAC GTGACCATAT CCAGCACCGT AGTTGCTACC	3540
	GAATACCACC ATGTCGCCAG GTTGTGCTAA GAAGTCCGGT GTATTTTGGT ATACAGTAGC	3600
	TAATCCGTCG AAGTTGTTAG CGAACGGAAT ATCTTTTGCA CCTAAACCTT TTAGAAGTAA	3660
45	TCCAAACAAA ACTTTCCAAC CAGCATTGGC ATAATCAAAG CATTGAAATC CATACCATAA	3720
	GTCCACATTG AATTGTTTTT CCTCAGAAGT TTTCAACCAC TCTATAAACT CATTTTTAGT	3780
	TAATTTTGCT TGCAATGTCG CCACCTCCAT GATGATACTC ATTCACATCA AAGCCAACAT	3840
50	CGTTAGAGGC GTCTGTGAAA GGTGTGTGATG TATCATATTC TTTTGGTGcT TTCGCGCTTA	3900
	ATTCCGGCGT TAACTACTG TCTTGTGATG ATTTCCACGT AACTTGTTGT TCTTCTTTTT	3960

55

	TTGGGTCAGT AATAACGCCA ATACCTGTAA GTAACGTGAG GATAGCGCCT ATAATTGCGC	4080
	TAGCTTGATT TAATTGAGTA GATAAATCTA ATCCGAATAA ATCCGTGACT TGCTTGATAA	4140
5	ATAGCAACAA TSCTCCAAC TAAACCACTTA GTACTGCTTT GTTTTTGAAT CTCAATTTC	4200
	AGTTAATATC CATTTGTTTG CTCCTTTTAT CCAAAATAAA AAAACGACTA AAAATTAGTC	4260
	GTTTAAATTT ATTCAATGGT CAATGTCGGA GATCCTGAAT AAACATCACT TATAGTGACG	4320
10	TACAACATCC CTGAAGGATT ACTAAAGTTG ATATTTTTAC TTGCAACTCC GCTATTGACT	4380
	CCTGATATTC CTAAATCACT TGACCCTAAA TTAGTTTGCG AAATCCTCAT TATACCGCTA	4440
	CGTACATTTT CTATTGTCAC CTGATAACTT TTATTGGGTT CAACTCCATT TATTGTCCAT	4500
15	TTTGCTGTTG ATTCTTCTAT GCTATCCGGA TATTTATTTT TAGGTAAGGG TTTTATTACA	4560
	AAAGATGAAG GCTTTTTTCCA TACTTGGATA TTTCCAGCAT ATACTTTTGT ATATTCTTCA	4620
20	CCTTCGTAAA TAAACTTCTT TACATTTTAA AAATTACCTT CCATAAAAAT CACCCTTTAA	4680
	TTAAATATAA CGTATTCGGG TCTTTTTGAT ATATATAGTT ATATTCAATT TCTGTTCTTG	4740
	TCCAAATTTT AACCGTCGGT TGAGATGCGC TTTTGTAGTT ATATAAATTA TCCGCTTGTT	4800
25	GTTTAGTAAA AGCTTGAGAT GACAAAACAT ACCGCTCGTC ATGATTATGA TTTTGTGGAG	4860
	CATATAAATC ATTTAGTGTT TGTTTGAATT CCTCAAAATC TTCTGTATTA ACTTTTGAGC	4920
	CAATCTGTTG CAATACACTT TCTGAAATAG AGTTGTTTTG TATTGCTTCT GCTAATTCTC	4980
30	TTAATGTGTT CATAGATTCA GGCGCGCTAT CAACTAGTTC AGCAATTTT GTATCCGTAT	5040
	ACGTTTTAGA GTCGTTGAGA GTTGATCTT TGATTTTTTC AACTTCTTGC AATTATTTT	5100
	CTAACCCTTC AACATTTGCG ATATTGATTT TGTCCAATAA CTCAGGTTCT GCTTTGATAT	5160
35	CTGTATCTTT ACCATCAATT TGCCACATTT TAGTGTCAGG ATTGATTGAT ACTACAGTAC	5220
	CGTTTACC GGGTGCCT TGTCTCCTT TTTTACCTGC TTCACCTTT GCTCCAGGTT	5280
	GTCCCGGTTT ACCTTTATCA CCTTTCGCAC CTTTAAATCT ACTTTCATTC TTTTCGATGT	5340
40	AAGAAATGAC ATCTTTATCT ATTTTCTCTT TAAAGTCTT GCTCAATAAA TCTGTGCGT	5400
	TATCTTTTAA AATTCTCGTA ATAGCATCAT CTACCAATT AACATCGATT TCTTTTGCTA	5460
45	CAGCAGATTC AATACCACTA TCAACGATAT TGAAAGAAAA GTTTGCGACA TGTATTTTT	5520
	CTTCTCTTT CTCTAAAAAC AGCTTACAGC GAACATAACC AGCGTGTG ATAACCTTTT	5580
	TAGGTATCTT GTAGGTAAGG AAACCTTTTA CAACATCGTC GATAATAAGG GGCTCATTTT	5640
50	TGAATATAGA GCCATCTTCC ATAACAAAT GTAATCTAGG TGTAAAGCCA TGTGCTTTTA	5700
	GATCGATACG ACCTTGTTTG TCATTGATAC CTATTCTTAT AGATGCTGTA TTTTCATCTT	5760

55



	CAACATCTTT TATTTTGTAC ATTTACACAC CTCTTTATTT ATATTTATCC CTTGTGAAGT	5880
	AGATACCTTT TAAGCCGATT TGTTTATATA ACTTAGCGAT TGTACTTGCT TGATGTTGGC	5940
5	ACCACTCTAT AGCAGTAGCG TATTGGTGGG TAGCTGGATT CTTAGGATTC CATCTAATTC	6000
	GGTACAATGT GTTTTGACCT TTATTGATGT AATCCTTTCT TACGAAGCTA GCACCGCCCA	6060
10	TGATTGCTTT TGCTGGAGAT GTCCAACCTT TATTCCTTGC AAACGTCATT GCGTAGTTAG	6120
	GATTGTTGTC GTAAGCGCCA ATGCCGAAGT AGTTGTATAC TCCATCTTTT CCGTTAGCGA	6180
	AGTTACTTGT TCCATATCCA CTTTCTAAGA AAGCATGCGC GATTAAATAA ATTTCAATTA	6240
15	TGTTGTGCTT TTTACAAGCT TCTGCGAAGC CTTTACCTTG ATTATTCAAT GTTCCCTTAC	6300
	CTTTAAGTAT CTTATTAAGT GCGCTAACTG AAACACCTTG ATACTTGCCT AAATTAAGCA	6360
	TTTGGTAGCA TTGTGTGTTA CTTTCCATA TACGCTTTAC ATTCATTGCT GAACTCGTTT	6420
20	GTGCTCGTGT AGCGTTAasc AACCCCAAGC ATTAGATTTT TTCGGGTTAC CTCTTGCCAT	6480
	TTGTTTATCC AGTGCTTGTT TGAATGTATA AGGACTCGTT TCTGTTATGA TCTGCGGTTG	6540
	TTTAGATGCC GAACCATTGT TGGCTGTTGG TGACGAGTCT CTTACATTAG CTATATCAGC	6600
25	GTTTTTATTA TCTACCATAA CTTTATTCT AGATTTTGTT ACTGTTGGCT TAGTTATAGA	6660
	ATTTAATAAT TTTTCTCTGT TTTTAAATAT ATTAAGTAAT GCCTTTTCTA ATGCTTCGTA	6720
	TTTATCTTTA GGAGGAACAC CGTTGTCAAT CATATTCCAA TTAACATGTT CCAACATTGA	6780
30	ACGCCAAATG CTGTCGTCTA CTTTAAATT TTCAATACTT AGAGGTATCT CATATTTGGC	6840
	CATCATATCT ACAGCTACAA CCATTGCGTG AATCTCATTA AAAATAAATT CATTTTACT	6900
	CGCACTATAA TCTTCACATA CGTCTATAAC TATATAATCA GGTTCATTAG GAACTTCAAA	6960
35	TACAGCTCTT CTAGGTGCCC AAATATTATG TCTATCAACA TAAAAGTGGG GATATTCTAC	7020
	ATCCTGTTTG TATTTCTTCC TACTGTTATA TAAACTTTCT ACCGAGCTCA TCGTTTGTGC	7080
40	GTTTCTAATC ATTATTCCTT TAGGTTTTTC GAGTCGTGCA TTACCTTCTA CTATAAAGTG	7140
	ATAAATATAT TCTGGATAAT TAACCTCTTG GCTAGAAATA GTGTACTTTA TAGTTGTTAC	7200
	ATCTTTCCAA ATTGGAACCT TTTTATTATT TTTTTCGTTA TCATCACTAT CATCTTCTGG	7260
45	TTTAGGTGCC GGTGTAGTTT TGTCTGGATG ATATGGTGGT CTAACAAAAT ATTTAACCCC	7320
	TCCACCTGST CCATCATGAT AAGAGTGTTC AATTTTATAA GGTGGACTTC CTGTTGCGTT	7380
	ATTTGTATAC CAGTTTTGAT CTACGCCATA CCAATAGTCT TTTGTGCATG GTCCCACTAC	7440
50	AATGTTTACA TGTCTGCCCC AACCAACAGT CCAAACACCC CAGTCGCCTG GTTGTGGTAC	7500
	AAAATCTTTT GTATTTCTAA TTATCTTGAA ATCTCTACCT CTATAATTGG ATTTTGTAGC	7560
55		

	TAAATCCCAG CATTGTGCTC CCATTCCAGA ACCAGGTACA TCAATAGCTA TTTTGTTTTT	7680
	AGCGATATAT AACGCCCAT T CAACCACTTC ACTAGCTGTG GGCTTTCTAT TTTTCGGATT	7740
5	AGGTAATCCC ATGTAIGCAC CTCATTTCAA TCAAAATAAA AAGCCAGTGC CGAAGCACTG	7800
	ACTCTTAACT GTTATTTACA TTTACCAAAC CAGAAGCAGC CCCAGAAGCT ATATCCTAAA	7860
	ATCCCTTTAA GCATGGTAAT CACCTCCTTT AAATACCAAA AACAGTTCTT AGTAAAGCTA	7920
10	TGACAATCGT ACTGAAGATA GTCCCTATCA AACCTAGAAT CCACATTTTT ATGTCTCTAA	7980
	TATTCTTGGC ATTCTTTTCT TTATCTTTT CATCTTCTAC CTTGTGCGGC TTAAATTCTT	8040
	CAAAATTTCT ATCTAATTG TCATAAATCT TTTCTTGCGC TCTAAGACTA TCTTCTATTC	8100
15	TGTCGAATTT TTCAAACATA GTCTTATCAT TTTCTTCTAA TCGCGTTAAA CGCCAATCTT	8160
	GTTTCATGTCG TTTGGTAAAT CCAAACATTA TGCCACCCAC TTTATTCAAA TTAAAAAGCC	8220
20	ACAAGCATT CACCTGTGAC TTTTCATCTT TTGTTTCTGG ATATTTTTCT CCAGTGATTA	8280
	AAGCGTATTC TTCTTTATCG ATTAAACCCT TGTCTACGTA CCACCTAATT TGCTCGTTTT	8340
	TATAGTAACC CCAAACATAA AAAGTTTTAA TGTCTTTAAA AGTTGGATAA ATCATCTTCA	8400
25	TTATTTAAAC GTCCCCCTCA GTACTTGTTT TGTTAGTTTT CAGTTCAGTC AACTGTTGTG	8460
	TTAACATAGC GTTTTGTTGA GCTAATTCCA TTGTTAATAC GTTTACTTGT GCCACCTGCA	8520
	TTTGCATACT CGCAACCATT CCGCGAAGTT CCTCATCACT TAAATCTGAC GCACTTTGTT	8580
30	GGTTTGATGC ATTCGGTACG TCTTCTTTTT CGAAATTGCT ATTGTATTTA ATTTGCGCGT	8640
	TAGTGAAAAC AAACCTTCTA GGTTCGAACT CTTCTTTAAA TTAAATAGGC ACATTGTTAT	8700
	CATCTACATC TAACTATTG CGTAAACCGC CAGTATTAAC GAATCCGATA ACTTCGTTTT	8760
35	TATCGTTTAC TGTGATTTTC ATTATTTCCA CCCATAATT TTAGTTATAG TAACTTTGTT	8820
	GGCAJTCGCT CCAGAACCTG ATGTTTTACC TAAATCAAAG TACACATCGT TATCTATTCT	8880
	TAAAGTAGTG CTACTTGTTT TGGATAGTAA GCACTCATAA ATACCGCCAC CGTTGCCGTC	8940
40	TGAGTCAACT ACATTGCTT TACTCAATTG AATCGCGTTA GGTAATGCGG TTAGTCCGAA	9000
	TCCCTCAATA ACGCCACCTG GATAAGTTCC ACTTACCAAC AAAATAGAAT AGTTTGTGTA	9060
45	CGGTTTCAGTT AGATTGATTG TTGTACCTAC ACCATTTGCG CCACCGTCGA ACAATACCGT	9120
	TGATTTATGT TCATTAGGAA CTGTCCACTG TTGCTCAAGT CTGCCGTTTG TGATTGATCG	9180
	TGTGTAAATC TTTTITAGAGT TATAAGGTGT GAAGTTAAAT AGCTTGTTTG TATCATCTTT	9240
50	AACGAATACC GATAAATAAC CCTCATAACT TTCAACGCTA CCTGGTAAAT CCGGCACTCT	9300
	TGTTGCATAG TAATTACCAG CAGTTAAATA TCCCAAATCG CCTTGCGCAT TATTTAAGTT	9360
55		

GAATTTATCA TCTACATACT GCTTAGCTTG ATTTAAAGCG TTGTTAGACG TTTCTTCAAC 9480  
 AAATTGCTTA GTTAAGTTTC CATCATTCTT TTTATAAAAC GGGTACCATG TGCCGTAGAT 9540  
 5 TTTGTATTTT GTGTACTCAT CGTTTGAATC GTCTGGGTAC CATGTTGCAC GAGCAGTATT 9600  
 ATTATCAACA ACATAAACAA CTAACACACC AGATTTGCTT GATGTATAAG TTGATTCATC 9660  
 GAACGAAGAA CCGTCATCAA CACCATCTTG TCCAGGCTTC TCTAACGTGC CTATATCCGT 9720  
 10 CTTTTCTGGC GCATCTGTTG CATTAGTAAT ATGAATAATC CTAGATGTGT TAACTGCGCT 9780  
 TAAAACGCTA TCTATGGACT GCTCATACGA TTCAATTGCT TTACCGTAAT CATCTGTAAG 9840  
 TTTAGACTTT TGCCAATTTC TTGTTGAATT ACCTTTAACA AGGTCAGCGC CATTGATTTG 9900  
 15 TTGTTCAACT TCGTTAACAC GTTCAAAAAT CGCTTGCTCT TTTTCAACTA TTTTATCGAA 9960  
 TTCAGCTGTA ACAGCTTGTG TTGCACTAGT TTGCGTCGCA GTAATAGCTT GTATAGCTTC 10020  
 GTTTTGCTTG ATTTTCGATTT GTTGAATGCC TTTTGTCGCA CTATCATTCA CTTTGTCTAT 10080  
 20 TAACGTTTGT GTATCAGCCA TATTTTGCTT TAATTGGTTA AAATCTTTAC CGACAGCTTC 10140  
 GATAGTATCT TGAATAGATT TGATATAAAC AAGCTTTGTT ATACCATCAA ACCCACTAAC 10200  
 TAAATCATTT TCAATATTGA AGCTAAATTG ACGTTCAACA ACAACATTAT TACTCCCGTT 10260  
 TTGTGTAAG AATGCCTGAG CATGCACCTT GCCTGAATGT TTTAAAAATT CATTCCGGTAT 10320  
 CACATACTGC AAACGCCCAT TAATTGCGTC TACTATCGTT AATTCGTCTG AAATATAAGC 10380  
 30 GCCTCTATCT ACGTTATAAT CATCGGTTTT TAAACGATA GATGTTTTAA CATGTTTACA 10440  
 ACTTATAGAT AAGGGTCTGT TATnCTTAGT 10470

## (2) INFORMATION FOR SEQ ID NO: 21:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3647 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

ATCAGATCTT GAGAATCGAG TTATTAAGTC TATCGAAGAC TTAATAAAA TCCAACCATT 60  
 45 CATGCCTACA CAAGATTTTG ATTTTAAAAC TAAAGAAATT CAATCAAACA TTTCTGAAGA 120  
 AAGATTTATC GAAATGATTC AGTATTTCAA AGAGAAAATA ACAGAAGGGG ATATGTTCCA 180  
 50 AGTTGTGCCA TCAAGAATTT ACAAATATGC GcATCATGCT AGTCAGCATT TAAATCAACT 240  
 TTCGTTTCAA CTGTATCAAA ATTTAAAACG ACAAACCCA AGTCCATATA TGTATTATCT 300

	TCAAATTGTA	ACAACTAATC	CTATTGCAGG	TACGATTCAA	CGTGGTGAGA	CGACACAAAT	420
	AGATAATGAG	AATATGAAAC	AACTACTTAA	TGATCCAAAA	GAATGCAGCG	AACATCGTAT	480
5	GCTAGTTGAT	TTAGGACGTA	ATGATATTCA	TAGAGTAACT	AAAATCGGTA	CCTCAAAAAT	540
	TACTAAATTA	ATGGTTATTG	AAAAATATGA	ACATGTTATG	CATATCGTAA	GTGAAGTCAC	600
	AGGTAAAATA	AATCAAAATT	TATCGCCAAT	GACAGTTATT	GCGAATTTAT	TACCAACAGG	660
10	TACCGTTTCA	GGTGCACCAA	AATTACGTGC	AATTGAAAGA	ATATATGAAC	AATATCCACA	720
	TAAACGGGGC	GTTTATAGTG	GTGGTGTGG	ATACATAAAT	TGTAATCATA	ACTTAGATTT	780
	TGCATTAGCA	ATTCGAACGA	TGATGATAGA	TGAGCAGTAT	ATCAACGTAG	AAGCTGGTTG	840
15	TGGCGTTGTA	TATGATTCTA	TTCTGAAAA	AGAACTGAAT	GAAACGAAAT	TGAAAGCTAA	900
	AAGCTTATG	GAGGTGAGCC	CATGATCTTA	GTTGTAGATA	ATTATGATTG	CTTTACATAT	960
20	AACCTAGTGG	ATATTGTTGC	TCAACATACT	GACGTCATTG	TTCAATACCC	TGATGATGAT	1020
	AATGTGCTGA	ATCAATCGGT	GGACCGCTGT	ATTATATCTC	CTGGTCCAGG	GCATCCATTA	1080
	GACGATCAAC	AGTTAATGAA	AATCATATCA	ACCTATCAAC	ACAAACCCAT	TTTAGGTATT	1140
25	TGTTTAGGGG	CTCAGGCACT	GACTTGTTAC	TACGGTGGAG	AAGTCATTAA	AGGCGACAAG	1200
	GTTATGCACG	GCAAAGTTGA	TACACTAAAG	GTTATATCGC	ATCATCAACA	TCTGTTATAT	1260
	CAAGATATAC	CAGAACAGTT	TTCAATTATG	AGATATCATT	CATTAAATAAG	TAACCCTGAC	1320
30	AATTTTCCAG	AAGAATTGAA	AATTACTGGA	CGTACCAAAG	ATTGTATACA	GTCATTGAG	1380
	CATAAAGAAA	GACCGCATT	TGGTATTTCAG	TACCATCCTG	AATCATTGTC	TACAGACTAT	1440
	GGTGTCAAAA	TAATTACAAA	TTTCATTAAT	CTAGTGAAGG	AAGGATGAAA	ACCATGACAT	1500
35	TACTAACAAG	AATAAAAACT	GAACTATAT	TACTTGAAAG	CGACATTAAA	GAGCTAATCG	1560
	ATATACTTAT	TTCTCCTAGT	ATTGGAACGT	ATATTAAATA	TGAATTACTT	AGTTCCTATT	1620
	CGGAGCGAGA	AATCCAACAA	CAAGAATTAA	CATATATTGT	ACGTAGCTTA	ATTAATACAA	1680
40	TGTATCCACA	TCAACCATGT	TATGAAGGGG	CTATGTGTGT	GTGCGGCACA	GGTGGTGACA	1740
	AGTCAAATAG	TTCAACATT	TCAACGACTG	TTGCTTTTGT	TGTAGCAAGT	GCTGGCGTAA	1800
45	AAGTTATAAA	ACATGGTAAT	AAAAGTATTA	CCTCAAAATTC	aGGTAGTACG	GATTTGTATA	1860
	ATCAAATGAA	CATACAAaCA	ACAACTGTTG	ATGATACACC	TAACCAATTA	AATGAnAAAG	1920
	ACCTTGTTAT	CATTGGTGCA	aCTGAATCAT	ATCCAATCAT	GAAGTATATG	CAACCAGTTA	1980
50	GAAAAATGAT	TGGAAAGCCT	ACAATATTAA	ACCTTGTTGG	TCCATTAAAT	AATCCATATC	2040
	ACTTAACGTA	TCAAATGGTA	GGCGTCTTTG	ATCCTACAAA	GTAAAGTTA	GTTGCTAAAA	2100

55

AAGCAACACT ATCTGGTGAT AATTGATAT ATGAATTGAC TGAAGATGGA GAAATCAAAA 2220  
 ATTACACATT AAATGCGACT GATTATGGTT TGAAACATGC GCCGAATAGT GATTTTAAAG 2280  
 5 GCGGTTCCACC TGAAGAAAAT TTAGCAATCT CCCTTAATAT CTTGAATGGT AAAGATCAGT 2340  
 CAAGTCGACG TGATGTTGTC TTAATAATG CGGGTTTAAG CCTTTATGTT GCAGAGAAA 2400  
 TGGATACCAT CGCAGAAGGC ATAGAACTTG CAACTACATT GATTGATAAT GGTGAAGCAT 2460  
 10 TGGAAAAATA CCATCAAATG AGAGGTGAAT AATATGACGA TTTTATCAGA AATTGTTAAA 2520  
 TATAAACAGT CACTTTTACA AAATGGCTAT TATCAAGACA AACTTAATAC CTTGAAAAGT 2580  
 GTGAAGATTC AGAATAAAAA ATCTTTTATA AACGCAATTG AGAAAGAACC AAAGCTAGCA 2640  
 15 ATTATTGCAG AAATTAAATC GAAGAGTCCT ACAGTTAATG ACTTACCTGA ACGAGATTTA 2700  
 TCGCAACAAA TCTCAGATTA TGACCAATAT GGTGCAAATG CCGTGTCAT TTTAACTGAT 2760  
 GAAAAGTACT TTGGTGGTAG TTTTGAAAGA TTACAAGCAT TGACGACAAA AACAAACATTA 2820  
 20 CCCGTATTAT GCAAAGACTT TATTATAGAC CCGCTTCAAA TTGATGTTGC TAAACAAGCT 2880  
 GGTGCATCTA TGATTTTATT GATCGTTAAC ATCTTATCTG ATAAACAATT GAAAGATTTA 2940  
 25 TATAACTACG CTATATCGCA AAATCTAGAA GTGTTAGTTG AAGTACATGA TCGCCATGAA 3000  
 TTAGAACGTG CCTATAAGGT TAATGCTAAA TTGATGGTG TAAATAACAG GGAATTAAAA 3060  
 CGATTTGTTA CAAATGTGGA ACATACAAAT ACTATTTTAG AAAATAAAAA AACAAATCAT 3120  
 30 TATTATATTT CTGAAAGTGG TATTCACGAT GCATCTGATG TAAGAAAAAT CTTGCATAGT 3180  
 GGTATCGATG GCTTACTAAT AGGTGAGGCG CTTATGCGTT GTGACAATCT ATCTGAATTT 3240  
 TTACCACAAC TGAAAATGCA AAAGGTGAAG TCATGATGAA ATTGAAATTT TGTGGCTTTA 3300  
 35 CATCAATAAA GGATGTTACA GCGGCCAGTC AATTACCTAT TGATGCGATA GGTTCATCC 3360  
 ATTATGAAAA AAGTAAAAGG CATCAAACAA TTACCCAAAT AAAAAAGTTA GCGTCTGCTG 3420  
 TTCCAAATCA TATCGATAAA GTATGTGTCA TGGTAAATCC TGATTTAACA ACAATTGAAC 3480  
 40 ACGTATTAAG CAATACGTCA ATTAACACAA TACAGTTACA CgGCACAGAA TCTATTGATT 3540  
 TTATACAGGA AATTAAAAAG AAATATTCAA GCATTAAAT CACTAAAGCT TTAGCTGCaG 3600  
 45 ATGgAAAAcm TwATCCCAAA caTtAAtnAA tnTTAgGGGG TCCGTGG 3647

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5966 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

	CcAcCTTGAC CACCTTTACG TGGAATCTTT TcMcCTkGAG CAACaTCGaT AATaTATATT	60
5	GAAAgTCAAC AAGTTCTGGA CTAAATGTTG CTGCTAAGTT ATCGCCACCA GATTCTATGA	120
	AAATTAGTTC TATATCGTCA TGACGTTCTA ATAATTCGTC TATTGCTGCA AAGTTCATAG	180
10	ATGCATCTTC ACGAATCGCA GTATGAGGAC ATCCACCAGT TTCAACACCA ATGATACGAC	240
	TTTCAGGTAG AACTCCTGAA TTTACTAATA TCTTTTCGTC TTCTTTTGTA TATATATCAT	300
	TTGTAATAAC GCCGATACTC ATTTCTTTTG AAAGACGTTT TACAACTTTT TCAATTAATT	360
15	GTGTTTACC TGCACCTACA GGACCACCAA TACCAATTTT AATCGGATTT GCCACAATTA	420
	TAACCTCCTA TGATATGAAA tTCTAACATT GaCGTTCTCA TCGGCCATTT GATTTAGTTC	480
	TAAACCAGGC GCTGTCATGC CAAAATCTGC TTCTTTTAAT TCGAAAATCT GCTTTCTTGT	540
20	TCCTTCTATA TAAGGAATCA TGTGAGTAAC TATCTTTTGA CCAGCAGTTT GTCCAAGTGG	600
	AATAGCACGA ACAGCATTTT GAGTTAAACT TGAAACATTT TGATATAAAT AGTAATCAAT	660
	AATCGTTTCA ATATCTACAC CTAAATGATG GCCTAGCATA GTAAAACAAA TAGCTGGATT	720
25	TnACTTTGCT TTCTTATCTT GCATTTGTTG ATGATACCAA GCAATCCATG GGCTATtATA	780
	AAGTTCTAAA GCCAATTTAA CCATGCGAGT CCCCATTGT kTTGCACCAA CACGTGTTTC	840
30	TTTAGGTAAG TTTTGtACAr ACATCAGTTT ATCTATGTGT AATACTTTTT GTGTATCATC	900
	ATTTTCCAAT GCATCATAAA CTAAaACGCAT GGCTAAACCA TCAGAATAGG TAAGTTGCTC	960
	TTGTAAAAAC ATTTTTAACC AAGCAATAAA AGTATGATCG TCATGAATTA TATTCGTTG	1020
35	AATATATGTT TCAAGACCAA ATGAATGACT GAAAGCACCT GTTGGAAGT GTGAATCACA	1080
	GAACTGAAAT AATCTTAAGT GTGTATGATC AATCATGAGA ATGCCCTATA TGTCTGAAAG	1140
	CCTTATTAAC TTTACGGTCT TCTCGAACAT ATGGGATGCC TAAACTTTTT AATAAATCTT	1200
40	CAACTAAATA ATCATATTGT ACTAGCATTT CAGTCTCTGT AAATTGTGCT GGCAAATGAC	1260
	GATTTCTTAA TTGATGGGCT ATATCTCCCA TTTCTTGCAA TGTCTTGGT TGAATCACTA	1320
45	AAAGATCTTC TGAATTAACA TCCACAATAA TCATATTATG GTCATCTGCG TATAAAATAT	1380
	CTCCATATTG TAAGTCAATA GGTGTTTTTA AACGAATGCC TATTTTCAGTG CCATGGTCTG	1440
	TAACGACTCT TTGAATACGT TTAACAAGAT CTGAATTTTC AAGGTATACT TTTTCGACGT	1500
50	GCTTTTGTTT TTCTGAATTT GACAAATTGG CAATATTGCC TTGGATTTCT TCAACAATCA	1560
	TTCTATGTTT CTCCTAGAAT AAGAAGTATC TTTGAGTTAA TGGTAACTCA GTTGCTGCAT	1620
55	TACTTGTAAT TTTTCTCCA TCTACATATA CTTCATATGT TTGTGGATCA ACGTCTAATT	1680

	GACGCACCAT GCGTTTTAAA TTAAATGCAC GATTGATACC ATTTTCATAA GCAGTTTTAG	1800
	ACACGAATGT CATTGACGTA CTTGTAAGGT TTCCGCCGTA TTGACCATAC ATTTTACGGT	1850
5	ACTTCATCGG TTCAGATGTA GGTATAGAAC CATTGTCATC GCCATTTACG GCAGAGTTAA	1920
	TTAATCOGCC CTTTACAAC TATTCAGGTT TAACCCCAA GAAAATTGGG TCCCATAGA	1980
10	CAATGTCAGC TAGTTTGCCC GGCTCGATAG ATCCTACATA TTCAGAAATA CCATGTGTAA	2040
	TTGCTGGGTT AATTGTATAT TTAGCGATAT AACGTTTGAT GCGATTATTA TCATTATGTT	2100
	CAAAATCACC ATCTAAAGGA CCACGTTGTT CTTTCATGCG ATGTGCTACT TGCCATGTTC	2160
15	GTGTAATTAC TTCACCTACA CGGCCCATTG CTTGTGAATC GGAACATAATC ATACTGAATA	2220
	CACCCATATC TTGCAGAACA TCTTCTGCTG CAATCGTTTTT TTTACGAATA CGTGAATCTG	2280
	CGAATGCGAT ATCTTCAGGA ATAGCCGCAT TTAATGGTG AGTAATCATT ACCATATCTA	2340
20	AATGTTTCATC TACAGTATTA TGTGTATAAG GCAAAGTTGG ATTTGTAGAT GAAGGTAAAA	2400
	TATTTGAAAA TGCAGCGGAT TTAATTAAAT CAGGCGCATG ACCGCCACCA GCACCTTCAG	2460
25	TATGGTACAT ATGAAGTACA CGGTCTTTAA CAGCAGCCAT TGTGTCTTCC ATAAATCCTG	2520
	CTTCATTTAA AGTATCTGCA TGTAATGCAA TTTGAACATC AAATTCATCA GCAACATCTA	2580
	ATGCATGACT CAAAGCAGAT GGTGTTGCAC CCCAGTCTTC ATGTACTTTT AATCCAATTG	2640
30	CTCCGGCATT GATTTGTTCA ATGAGTGCAG TTGGATTTGT TGCTTGCTCT TTACCTGTAA	2700
	AACCGACATT AATCGGTAAA CATTTCGGCAG CTTCTAACAT TCTATGAATA TGCCATGGAC	2760
	CTGGAGTTAC AGTTGTTGCT TTAGAACCTT CTGAAGCACC AGTACCACCA CCAATATGAG	2820
35	TCGTAATACC ACTTTCTAAT GCGACCTCTG CTTGTTTCAGG ATTAATAAAA TGAACATGAG	2880
	TATCAATACC ACCAGCAGTG ACGATTTTAC CTTCAGCGGC AATGATATCT GTTGTGTAAC	2940
	CTATAATAAT GTCGACATTA TCCATTATAT CTGGGTTGCC GGCATTACCT ATGGCGAAAA	3000
40	TATAACCATT TTAAATGCCT ATATCAGCTT TAACCACTTT ATCGTAATCG ATAATAACGG	3060
	CATTAGAAAT GACAAGGTCT GCAACGTTCA CGTCATCACG TGTTACACGA GGATTTTGCG	3120
45	CCATACCGTC TCTAATAGAT TTACCACCAC CAAAAGTAGC TTCTTCACCA TAAACCGCAT	3180
	AGTCTTTTTT TATTTGAGCA AATAGATTCTG TATCACCTAA ACGAATGGAA TCTCCAACAG	3240
	TTGGACCGTA TAAGCTCGTA TATTGATTTT GCGTCATTTT AAAGCTCATG ATCTTTTTTC	3300
50	TCCTTTTTTA TTCACGTTTT CAGCACCGTT ATCTCCGAAT ACACCTGCAT ATTCATCATT	3360
	TTCATCAGTT GGGCGATAGA CACGTGACTC ATCGATAGGA CCATTGACCA TACCACGAAA	3420
55	ACCAAAAATT TTACGTTTGC CAGCATATTC AACTAATTGA ACTTCTTTTT TATCCCCAGG	3480

	TTGAAATCT AATGCTGCAT TTGCTTCATA AAAATGAAAA TGTGAGCCCA CTTGAATTGG	3600
	TCGATCTCCT GTATTTTCAA CTTGATAAC TGTTCAGGA TGATGGTTAT TAATTTCAAC	3660
5	CTCTGIACTT TTTGTAATAA TTTCTCCTGG TATCATTGA CTGCCTCCTT TAAACAATAG	3720
	GGTGATGTAC TGTGATTAAC TTAGTACCAT CGGGGAACGT AGCCTCGATT TCGATATCTG	3780
10	TAATCATGTG TTCGACACCA TCCATGACAT CTTCTTTGTT TAGAATTTGT CTACCATAAC	3840
	TCATTAACTC TGCAACGGTC TTACCATCGC GTGCACCTTC TAATAATTCA TCGCTGATTA	3900
	AAGCTAATGC CTCAGGATGA TTTAGTTTCA AACCACGTGC TTTACGACGA CGTGCAACTT	3960
15	CCGCCGCCAC TACAATCATT AATTTGTCTT GCTCTCGTTG TGTAAAATGC AAATTAAAAC	4020
	CCCCAATTTC ATATTAGATA CaATTTACAA AATTTATATT AATCCTAATT GTTGTGATAA	4080
	ACAAGTAATA TACAAAGTTC AATGTGTAAT TAGAAAATTA TATTTTITAGC ATATCCGATA	4140
20	TTGAAGCAAA CAATCTAATC GAAAACAAAT AGTGGAATAT ATTTATGTAA AAACCAAAAT	4200
	AGTTTTTAAT ATAACTTTTC ATAGAATAGT AGTATATTAA TGAGTAATGA TTCAAAGGAA	4260
	AGGTGAAAGA TTTGAAGATA ATAGATGTGC TTTTGAAAAA TATATCTCAG GTTGTGTTAA	4320
25	TTAGTAATAA ATGGACAGGA TTATTTATCT TAATAGGATT ATTTGTAGCC GATTGGACAA	4380
	TTGGATTAGC GGCTATTGTA GGTAGCATCA TCGCCTATAC TTTTGCGCGT TTTATAAAAT	4440
30	ATAGTGAGGC AGAGATTAAT GATGGGTTAG CTGGATTAA TCCAGTGCTA ACTGCCATTG	4500
	CGTTAACAAT CTTTTTAGAT AAGTCAGGAT TAGATATTGT TATAACAATG ATAGCAACTT	4560
	TATTAACGTT ACCAGTTGCT GCTGCAGTCA GAGAAGTTT AAGACCATAT AAAGTTCCGA	4620
35	TGCTGACGAT GCCTTTTGTC ATTGTGACTT GGTTTACAAT TTTACTTTCA GGACAGGTTA	4680
	AATTTGTAGA TACATCGTTA AAGTTAATGC CTCAAACAT TGAAACGGTT AATTTTAGCA	4740
	ACAATGATAG AATaCATTTT ATTGAGTCAT TATTTGAAGG ATTGAGTCAA GTATTTATCG	4800
40	AAGCGAGTGT AATTGGTGGC GTATGTATTT TAATCGGCAT ATTGATAGCA TCAAGAAAAG	4860
	CAACACTCTT AGCTGTTATA GCTAGTTTGT TAAGCTTTAT CATTGTAGCT CTATTAGGTG	4920
45	GTAATTATGA TGATATTAAT CAGGGATTAT TCGGTTATAA CTTTGTATTA ATGGCAATCG	4980
	CACTAGGATA TACATTTAAA ACAGCGATTA ACCCTTATAT TTCGACTTTT TTAGGTGTGT	5040
	TATTAACAGT AGTGGTGCAA CTAGGTACAA CAACATTGCT TGAACCGTTT GGCTTACCTG	5100
50	CATTAACATT GCCATTTATT ATCGTGACAT GGATTTTATT ATTTGCTGGT ATTAAACATG	5160
	ACAAAGTAGA TGCTTGATAG TTAAATCAAA CCTAATATTG TTTGAATATC ACCTTAACT	5220
55	ATACAGCGAA TTGTATAGTT TAAGGTGTAT TTTTATGGAT AAAATTAAGT GCATACTTAA	5280



GTGTAAACT AGGAATAAAT AATTATATT GTGTGTTGTG TGGGGTGAAT AATATGAATG 5400  
 ATATGGATAA TTCCTTTTTA ATAACAACGG AAATTCAAAG AAAATGGATT GAAAAATTCA 5460  
 5 AAGTAATTAG AGATACATTT AAGGCTAAAG CTGAATATAA TGATCAACAT AGCCAATTTTC 5520  
 CATATAAAAA TATTGAATGG TTAATTAAAG AAGGTTATGG AAAATTAACG TTACCAAAAAG 5580  
 CATATGGTGG TGAAGGTGCG ACCATAGAAG ACATGGTTAT TTTGCAATCA TTTTATAGGCG 5640  
 10 AACTTGATGG TGCCACAGCA TTATCTATTG GTTGGCATGT GAGTGTCTGA GGACAAATTT 5700  
 ATGAACAGAA ATTATGGTCT CAAGATATGT TGGAGCAATT TGCTGTTGAA ATTAATAATG 5760  
 GTGCATTAGT TAATAGAGCA GTTAGTGAAG CTGAAATGGG TAGTCCAACA AGAGGGGGAA 5820  
 15 GACCAAGTAC ACATGCTGTT AAAGCTGATG ATGGGTATAT TTAAATGGT GTGAAGACAT 5880  
 ATACATCAAT GAGTAAAGCA CTAACACATA TTATTGTTGC TGCTTATATA GAAGAATTAG 5940  
 20 AAAGTGTTGG TTTTTCTTA GTAGAC 5966

## (2) INFORMATION FOR SEQ ID NO: 23:

## (i) SEQUENCE CHARACTERISTICS:

25 (A) LENGTH: 17310 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

CTGTGTCATC GCGAAATAGT TAGGGTCATT CATTAACTCT TTTGAACGTA TTTTCATCAA 60  
 35 ATATAACAAT TTCATTAGTA AAGGGGACTT GTTCAAACCA GCTATAATAC AAAATAGACC 120  
 TATAGTCACA CTGCTTATAA TATAAGAGGT AACGATCACT TTTTGTCTAT TACCTAACTT 180  
 AAAGATGATC ATCCCTAAAT AGAAATAAAT GACTACAAAT GCATATTTAA CTGTAGATGC 240  
 40 AAGAACTTCC TTAACCGTAA TAAATATCAA ATCATCAAAA AATaGCaAC AArGCGTAAT 300  
 AATCATACGA TATGTATACA AAATAATGA<sub>m</sub> AAAGTGT<sub>m</sub>AA AAATGATTTG CCTTTAATAA 360  
 ATGGTTAGCG AAAACAGTA AATAAACTAA TATTAGTAAT GTGATAAAGT CAGCTATAGA 420  
 45 AACATTCAAC CCGGCAATAA CCGAAGATTG CTGAATAAAA ACCGCTAAAC CGATAAGTAA 480  
 CAATGTTAGT AATTTACTAT TGTGTTGATT TTCCATTATA AACGTCTTCC ACTTCTTTAA 540  
 TCATTTTCTC CTCAGTAAAA CATTCTAAAT AACGTTTTCT AGATTGATTA CTCATTTTGA 600  
 50 TGTAATCACT GTCTATTAAA TATTTTCCCA GGACTTTAGC AATAGTTTCG GGTGTTGTG 660  
 TCATCATACA TATACCATTA TCAGCTACTA ATTCTGAAAT ACCGCCAACA TGAATGGCTA 720  
 55

	TTATTAAAT AAACGTATCG TATTGTGATA ATAAATGACT CGCATTAAATG ACATTGCCCA	840
	AAAATGTGAC ATCATTTTCT AACCAGCTT GTACAACTTG TTGCTGACAA TCATTTAATG	900
5	TAGGTCCATC GCCTATAAAT GTAAAATGCG CATGATTACT GTTATGTAAT TTCAATATCT	960
	CTATTGCCGC GATTAGATTT TGTGGCAATT TTGGATAAGC AAATCTTGCA ATCATAACAA	1020
10	ATTGATGCTT TGTCGGGGCA TTAATCTGTA AATCTTGTTT ATTAGGCAAC ATTCCAATA	1080
	CTTCGCCAAT ATTGTTATGT GATTGGCTTT TTAGCGTTTG CTTAACAGCG GGAACATCTG	1140
	CAATACCAAT ATGTATTGTG GTTAATTTCA ATCGATTAAA TCGATATTTT AACGCTAACT	1200
15	GTTTATCGAA ATCTGAAACA CAAATAATGC TATCTGTAAT AAGTGACATT AATTTTTCGA	1260
	TAACTAAATA TAGAAATTTT TTAGCTGGTT TAACACCCTC TGTAAGGCC CATCCATGTG	1320
	CAGTAAAAAC TATACGTGTG TCTTTCGATT TCGAAATGAa CTtCGCAATT CGTCCGACCG	1380
20	TtCCAGCTTT GGAAGAATGT AAATGGATAA CATCAGGTTT AATTTTCGAG AATAACTGTG	1440
	CTAACACTTT GACAGCTAAA ATATCTTGTT TAAAGTCAAT TGGACCTACT AAATGTTCTGA	1500
	TAATAATTAC ATTAACCTCT GCATCTAGTT GTTCAATCAT TGGTCCATGA TTGCCTACAA	1560
25	TGACATAAAC ATCATTGTGT ACGCAAAAAT GGTTGGCGAG TTGAATGAGA TGTGTTTGTG	1620
	CACCACCATT GTCTGCTTTA GTAATACAAT ATATAATTTT CAACTGTTAC AAACCCCTTT	1680
30	AATGCTATAC TTTCAATTC TTAACATGGC TATCTCATCA GATGAATAGT ATTTATAGCC	1740
	ATGCAAAATCA ATGATGGCAC ATATTTCTTA ATGCCATTTG ATACTGTCTC AAGGGATTCC	1800
	TCGTTATACT GTAACAATTG GTCACAATCT TTAATAATA ACTTTTATTT GAACTTATTA	1860
35	AGTAAATTAA GACTACCTTG AGCCTTCCCC TGTAATAACA ACCATCAATG TTCTAATTGA	1920
	TATATATAGT TCCATCATTA AACTACCTTT ATGTATATAT TTCATGTCAT ATTTCACTTT	1980
	TTGTTGCGGT GTTAAGTCAT ATCCACCTTG AATTTGCGCA AGTCCTGTTA ACCCTGGTGT	2040
40	AACAAGACAT CTTTGCTCGA AACCTATCAC TTCTGAACTA AATAATTCTA CAAATTCCGG	2100
	ACGTTCCGGG CGTGGTCCAA TAAACTCAT TTCCCCTTTA ACAACATTAA TTAGTTGTGG	2160
45	TAATTCATCA ATGCGTGTTT TACGAATAAA CTCCCGACA TTTGTTATAC GATCATCATC	2220
	TTTATCAGCC CATTGCGCAC CGTTTTTCTC TCGGTTTTTG CACATCGAAC GTAATTTGTA	2280
	TATTTTAATT AATTTACCCA TCTTCCCAAC TCTAACCTGA CTATAAATAG GGTTCCTGG	2340
50	CGAATCTATG ACGATAGCAA TGGCGAATAT AACCATAATC GGTAAAGTTA AAAATAATAA	2400
	AACAATGCTT AAAATTAAGT CAATCGCACG TTTAATTGGG TAATAGCTTT TTCTCACTTC	2460
55	TTCTAGTTTG TCTAATTTTC TTTGATAGGC ATAACCCTTA TTATTATGGA CAGCTTCAAT	2520

	AATTAAAGTA ATCCTTTTAAA CCTGTTTCTA CTGTATATTT AGGAACAAAT CCTAATGCCT	2640
	TTAAGTTAGA AATATCTGCA TAAGAATGCT TAATACTCTC TTTTCGTGCT TCTTTAAATT	2700
5	CATGCTCGAC TGATTTTCCA TATAATTCAC CAATAATACG ATAAACCTCT AATAAATTAG	2760
	TAAAAGTGCC TGTACCAATG TTATAACCGT GTCCAATTGC ATCTTTGTGT TCCATAATTA	2820
10	AGCGTACAGA TTGAACAACA TCATATACAT ATACAAAATC TCTAGTTTGC AGTCCGTCAC	2880
	CAAAAAATGT AAATGGCTTG TTATGCTCAA ATGAATCGAA CATCTTTGAA ATCACACCTG	2940
	AATATTGTGA CTTAGGATCC TGTCTTGGCC CAAATACATT AAAAAATTTA ACAACCGCTG	3000
15	TTGGTATGTT ATATAACGAA CAATAATTTA ATGTGCTCCG TTCGCCGTAA TATTTATCTA	3060
	TTGCATATGG TGATAATGGT AAGATTAATG ATTGATCACT TTTAGGCAAA TCAGGAAGAT	3120
	CACCATAAAC AGCTGCTGAC GAAGCAAAGA TAAACGTTT TATATGATTA TTATATTTTT	3180
20	TAATGATTTT TAACAATCTT AATGTTGCTA CGACGTTTAT TTCTTGAGAT AAGATAGGTT	3240
	TCTCAACCGA CTCAGCAACA CTAACCTAATG CTGCTAAATG AATAACATAA TCAAATTGAT	3300
	ATGTCTTCAT GATTTGTTCA ACTGCATCAT ATTCACGAAT ATCTAATTCA AACACATGAT	3360
25	CGTCAGCCAA ACTTTTAATA TTTTCTCGTT TACCTGTTCT ATAGTTATCT AGAACATAAA	3420
	CATCATAATC TTGTTGTAAA TCATCTACTA AATGCGACCC AATAAAACCA GCCCCACCAG	3480
30	TTATCAAAAC TCTTTCCAAA TCTTCCACCT CATTATATACA TTAAAAATAT ATCATAAAAA	3540
	CATAAAGTAT TGTAAGCTTT TTATCGATAT TTTTATTTA TAAAAATAAA ATGAGATAAC	3600
	TTTGTGAATT TTTATTGAGA TAAATTAGAT AGTGGTGTTT TTGTGATGTT TTATAATATC	3660
35	TTGGGTGTGT TAATACTAAT AATGCTTTCA ACTGATGCAT TAGACTGTGA CATCATAACT	3720
	CACTTAAGAA CTTCGCTTAT TAATTTTCTA CCAATACACT CCCTTCTAAG TGCACTAAAA	3780
	AATCCTTACT GCTAAGTGAT TAACTTAAC AATAAGGATT TATTTATCAT TAGTGGATGA	3840
40	TTATTAACGG AATCTCATAC CACCATCTAC AATAATTGTT TGTCCAGTAA TGTAATCAGA	3900
	GTCTTTACCA GCTAAGAAGC TCACTACATT TGAAACATCT TCTGGTTGAG AAACCTCGCC	3960
45	CAAAGCAATC TGAATTGTAA ATTGTTCCCA ACCCATGCT TCAGGTTTAC CTGCTTCTTC	4020
	GGCTGTTGCC ACTGCGATAC TTCCATCAT TGGTGTGTTGA ACGATACCAG GTGCGAATGC	4080
	ATTCACAGTA ATACCTTCAG ACGCTAAATC TTGTGCGGCT ACTGTGTGTA AACCTCGCAC	4140
50	TGCGAATTTT GTAAGTCAAT ATAAAGACAA GCCTGGGTGA CCCTCAACGC CTGCTTGAGA	4200
	TGTTGCATTG ATAATTTTAC CGCCATGATT GAATTTTTTA AATTGTTTAT GTGCGGCTTG	4260
55	AATACCCCAT AGCACACCTG CAACGTTTAC GCCATATACT GTTTTAAACT GTTCTTCAGT	4320

	GCCAAATTGC GCGGCAGTTT GTCTTAcTGC GTTAAATACA TCATCACGGT TTGATACATC	4440
	TGCTTTGATA GCAATAGCTT TTGTACCATC ACTTGATAAT TTAAGTGCAG CTGCTTTTGC	4500
5	CCCTTCTTCA TTGAAATCAA CAACTGCTAC TTTGAAACCA TCTTCCACTA AACGTTCTGC	4560
	AAATTTAAAA CCAATCCCTT GTGcTCCGCC AGTTACTAAT GCTACTTTGT TGTTTGT CAT	4620
10	AAAGATCACT CCTCAAATTT CTTTCCTTTA ATTACATTTT ACTCCTCTTC ATTTGAATAG	4680
	TACAACAAAG GTAGCTCCAT TTAACAAAAT ATTACAGATAT TTAAGGTATA GTTAAACGCA	4740
	CTACCATTAG TGATTGGCAA TCGCTTTAAA TGTCGTTTAA AAAGTTCTTA TGTTGAATAT	4800
15	TATTTTTTTA AGTCTCTCGA TTAGTTTGTC ATCAATCTTT TTTGAGACA TGGTCTTTTG	4860
	ATTCAATAGG CGGTTCCGTG TTATCACTGA CAACTTTAGT TGTAGCTTCA TCTTTATGTA	4920
	TTTCTTCGTT AAATCCTTCA AGGTTTTTAG TCGTGGGATT TTTAACCTCA GGATGTTCCA	4980
20	TCATGTCTTG ACTATCAAGT TCCTTTTTAC ACGTGTCTTT ATGTGATGCT TGATTTGCGT	5040
	TCCCTTTACT TTTTTGAATA GTGGTAGTAT CTGCTGCAGC TACTAATTTT TTTCTACCTA	5100
	AAATAGATAT GGCTGAAACA AACCAGAGTA TTGCAGATAC AAAGTTGCAT AATACTAAAG	5160
25	CGATAATAGC CAATACAATT AATATGACAC CTTTTGAAAT CCTTTCTTTA AATAAGTCAG	5220
	ATGCCAATAC GATGACAGGT ACGATTGAAA GTATAATTAC AAATATAGAA ATTATTGCCG	5280
30	ATATAACTAT TGTTACTATT AAATAATCAG CTCTGCTACC TGATAATAAA TAGAAAAGGC	5340
	CGAAAATTAG TCCATAGCAA ATTACAAACC CACATAAAGT TATAGCCATG AGTACTATAT	5400
	AAGCTATTTG AAAATATAAA CCTATCTTTA TGAATGATTT TTCTACATTT TTTTCCATGT	5460
35	CTATTCCTCA TTTATTTAAA ATTTATACTT TACCTTAAAT ATTCTCTTTA TTCTTTAGTG	5520
	ATTTTATCTT TAGATTCAAA TTGATTCTCT GTACTTTCAA TATCAACTTT TTCATTTTCG	5580
	TCTGTTCGATT CATCTTTTGA GTATTTATTC CAAATCAGCA AAATACCACC AATCAGCCAT	5640
40	AAAATTGACG AAAGGAAATT ATATAAACAC AGTGCAATAA TAGCATAAAC AATAAAAAGT	5700
	GCACCTCCGA TTACAGAGTA ACTTTCCATA TAAATCGCAG TAAAGATGGT TGGTAAAACA	5760
45	GTGAAAAGAG CCAATATTAA TCCTAATAAA AAAATTGTTT CGTAATCAGA TCCTCCAGCA	5820
	ATATTAATAG ATATCATCCT AACAAAAACG AACTATAAAT ATATTGAGC TACGATGCCT	5880
	ATCCAAATTG CTATTTTTC TATAATTGAG CTCATACTCA TTCCCATTT ATTTAAAAAT	5940
50	TATACTTTAC CTTAATATAC CTTATTTTAT TTAATTTTAA TATGCAAAAT ACAAAAATGG	6000
	AGAACTTCAA TATTTATAAA ATATCAAAAAG TTCTCCACAC TATATTGTTT TATTATATT	6060
55	TCGCTATCAA TACGCTAAAT CATCATATTT CCCTCAACAT CACAGTAAAA CTATTGCTCC	6120

	TTCCAATTGC GCAGTTGTTT AACATCATCA TCTTGTTTAA GTAATGCCAG TGGTACTTGA	6240
	AGATTAAGAC ATCGTCCTGA AATATTAAAG CGTGTCACAC CTGCTGGCAC AGTTTCCCCT	6300
5	TTATGAACAA CCGCTTCAAT TTCCTTATAA CTCAATGGCT GATACTTCAT GAGTACATCT	6360
	TGTTGAGAAA GACAAGGATA TGTACCTTGT GCAATTCTCT CTACAGAACA ACAACCACTA	6420
10	TAACTTGCGA CAACCTTTTC CCATACTTGA AAATGTGCTT CGCCTAAATC TTTTGTATAC	6480
	AAATATTGTT CTGTATCACC ATGACACATT GTAATAAATG GCGCTTCTTG TCTTGTCTCA	6540
	GTAGTCCATG GCAAGCGATG TTCTTGTTGT AACGTTTCCC ACCACACACC AAATGGAACCT	6600
15	TTATGTTGCC ATGTACTAAT TGAATATTGT GTTTCATGGA TTTCTTGAC TGGAACTTTC	6660
	TTACATCCTA ACGCTTTCAA ACTTGTATAC CGATGCACAC CATCTATAAC CATATATCTA	6720
	CCATGTTGCA TCGCTGTCAC TAAAATAGGA TGACGTATAA AATCATCTGC TTCAATACTA	6780
20	CTTTTCGTTT TTTCCAATCT TAAAGGTTTC AATGTTTCGT GAAGATCAAT CTTATCTACT	6840
	GGTACCAATT TTAAATGTTT ATGAATATGA TTCAATAGTT ATTCATCCTC CTTTGTGTTT	6900
25	GTTAAATAAA TAAATTCAGG ATGTGGATGG CTTAAGAAAT CGTGATGTGA AATAGACCAT	6960
	CCGTATGCAC CTGCATATTT GAAAACAATA ACGTCGCCTG TACTGATTGC GTCTATCTGT	7020
	ACTTCTCTAG CAAAGACATC TTTCCGTGTA CATAATTGAC CGACTAACGT TGTGTCCTGT	7080
30	CTCGAAATTG AAACCTTTTC AAATGAATAT GGATTGTCCT TATAGCGATA AATGTCAAAA	7140
	GGATGGTTAT GTTGCCAAGA TACCGGCAGT CTAAATTGTT GCGTACCTCC TCTTAATATG	7200
	GCATACCAAG CACCATGTAC TTTCTTAATG TCTAGCACTT CTGTCACATA GTAACCAATA	7260
35	TGTGCCACAA TAAAGCGCCC ACATTCAAAG TTCAATGTCA CATCTTCCAT TTCTTGCTCA	7320
	ACGATAAGTG TTTTAAAACG TTCTACAAAA TTATCCCATT CAAATTGGTT AGTTAAATCT	7380
	GCATAGTTAA CGCCTATGCC ACCACCAAGA TTGATATGTT TGAGTGGAAT TCGATGTTTT	7440
40	TCAGACCATG CCTTTGCTTT TTTAAAATAA AGTTTCACTA CATCGACATG TAAATTCGAG	7500
	TCTAAATTGT TAGAAATAGA ATGAAAATGA AATCCATCTA GATGAATCTT TGGCATTGCC	7560
45	AGCGCAGcTT cAATGACATC ATCAACTTCG TCTTCAGAAA TACCAAATTG TGTGGGCGT	7620
	CCTGCCATAT GCAACGTTGC ATTGGGAAAT GGTCTGCTA AATTAACACG CAATAAAATG	7680
	TGTTGTGTCT TATCTTCATC TTCTAAGATG GCATTTAGCC GTTGTAATTC ATGCATACTT	7740
50	TCAACATGAA TACGCTGAAC ACCTTCACTT ACTGCATATC TTAGTTCCTC GTCTGTCTTA	7800
	CCAGGGCCAC CAAAAATAAT ATGATTTGCT GGTTTAAAAG CAAGACCTTT TGCTATTTCA	7860
55	CCTTGAGATG CAACTTCGAA TCCTTCAACA TACTGACTAA TTGTATCTAG GATTTTTCGT	7920

	TGTTGCAAAT GATGTTCCAG TCCGACTAAA TCATAGATAT AATGACAAAC TGGATGAGAT	8040
	TGTGCTTTTA ATTGTTCAAT AACAGGTTGA ACTATACGCA TTAGCCTTCA TCCCCTTTCT	8100
5	GTTTAGACGT CGCTAGAGAT GCACTTAAAT GGCGATATAT TTTTCCGCGA TCATCACCTA	8160
	AAATAAATGT TTGTACACCT TGTGCCTGCC ATTTTGCAAT ATCTTCATCT TCACGTGGTA	8220
10	ATGCACAAAA ATGTTTACCA TGTGCATTCA CAACTTCAAA AATATGTTGA ACATGTGATG	8280
	TTACTTGATC ATCACGCGTT TGCCATGGTA TGCCAAGTGA CTGCGATAAA TCTGCGGCAC	8340
	CTTCGACTAT CATGTCTAAA CCTTCGACTT GTGCTATATC GTCAATGGCC ATAACCCCTT	8400
15	CAACATCTTC TATCATGGCA ATCACCATAA TATGCTCATT AGCCATCTCC ATTGCATCAA	8460
	GTAATGGTGT ACGTCCAAAT CTTGCCATGC GACCACCATT CAAACTTCTT AATCCTTGCG	8520
	GGTAATAACG ACTTAATTC ACAATATGCT CAACTGTCTC ACGATCTTTA ACGTGTGGCA	8580
20	CAATAATACC TCTCGCACCC ATATCCAACA CTTTAATGAT ATCTCTATCT ATCACTGCAG	8640
	TGACACGTAC AATTGGTATA ATATGCGCTG CTTCAGCTGC ACGAATTAAA TGCCTAGTG	8700
	TCTCATCATT AATCGCCACG TGTTCTGTAT CAATCACAAC AAAGTCATAC CCGCTTGCTG	8760
25	CGATAACCTC GATCATCAAT GGGTCCGGTA TAGAATTAAA AATGCCATAA ACTGAATCAC	8820
	CATTGTTTAA TCTATGTTTC AGAGATAGTT GTTGTCATCAT TGATACCTCC TACACCTAAT	8880
30	GGATTTGTAA CATGATGAAT TCTTAACTCG GAGTCACTTA ATAATCGACG TGTCGTTAAC	8940
	TTTTCAACTT GAATCGTAGG TTCAAACAAA TCGAAATGTT GATAGTTATT CAACTCTGGA	9000
	AATGCTTCTT GATACGCCTC GATGATGCCT TTAACCCATT GCCATTGCAG CTCCTCATCG	9060
35	ATACCATATT GCTTTTCAAT AAATAAGATG ATTTCGGCGA TATTAATAAA GAAAAATGCA	9120
	TCATGTAAAA AGTCGCGTAC TAAACGTTCC TCATCTGTTT CAATAAATGA ATTACTATTC	9180
	ACTTTTATAT GTGCTTCTGG CATTGGCTTT AATGTCAGGT GTGAAGCAGC TTCACTTAAA	9240
40	TGctCACGCT TAAAACGAAC ACCATCATGG AAATCTTTTA AGGCAATACG TGTAGGCCAA	9300
	CCATTTTCAT GAATGAGCAT CATATTTTGT GCATGCGATT CAAAGGCAAT ACCGTGATAA	9360
45	TAAAGCATAT GAATCATTGG ACGAATCGCT ACAGCTAAAA ATTGCTTTGT CCAAGCTTCA	9420
	GAACCATATT GTTTAATCCA ATTTTCAATG AATGGTACAC CATCCTTATC ACTTGATAA	9480
	AGTGCATTAA ATGGTATCGC ATCCTCTTCA TCGATTAAAC TATGATATAT ATTTTCACGC	9540
50	CATATAACAC CTAACGCACC ATAAACTTGA GTTTGTTTAT AAGGCGAAAG TTGTGTATTT	9600
	AAATAAGACT GTCCTAAGAC TTCCCCTAGA AAAACTGTCT TTAATTTCATC TTTTAAATAC	9660
55	ATATCTTGTT GCTGTATCTG CTTTAACCAA TCCGTAATTT GCGCTGCATT TTCAATTGTA	9720

## EP 0 786 519 A2

TATTTTGTCTG TGTCTATTGG CGACATCGTA CGAATCGATT GTTGAGGGTG ATATAGCTCA 9840  
 TCACTTTCCC CTAACCATAG TACTGTGCCA TTAAGCCTTT CTTGAGCCAA ATCAACTTGG 9900  
 5 ATGACATGTT CAAACTGCCA TGGGTGTACA GGTATCATCT CAACATCATT TACATGTTTG 9960  
 CCAGATGCTT CAATTTGCTG TACAAAATGT TCATAAGTCT TATCGCCAAC TTGTTGACGT 10020  
 AACATTTTCTG TAACTACAAC ATTTCTTGAT ACCGTCGTTT CTACTTTATC TTTGTGCGATA 10080  
 10 GCTAACCCTT GCAGTTTAAAC GTTTGGTACA AAATCAGGAC CAAATTTCAA ATTATCACTC 10140  
 AACGTAAATC CTAAACGTGA TTTGTAACCT GGATGATACT GATGCCCTTC CATCGCATAA 10200  
 15 AATTCATAGT CGTTAAATGT CTCAGGTGTT GCTGGTGGGT TTGATTCTCG ATACTGCATA 10260  
 CTTTGCCTAT CTTTAAATTC TGTCTGTAAT AACTCGACAA TAAATTGTTT TAGCTTTTCA 10320  
 TCATTTTTAG GAAATGTAAA TACAACCTCT CTCAATAATT GTGTATAGTC TGTGTTGTA 10380  
 20 TCTGCCTCAT CTCCTACGAC ACGCTCAATT GGTGATGTGA TACGTATACG ATCAAAGCTA 10440  
 TGTGTCTTTT CAGCAGTAAA ACGATACTCT GAATCATGTC CTTCTATTGT AAAATGACCG 10500  
 ACACCGTCTT GATATGACGC TTTATACACA ACAATATTCT CATAAATAAG TGATGATACC 10560  
 25 AGTTGGTGCA TCACTCTAGT CTTTACACGA TTAAGAATTG TTTGATTAC AATACGATAC 10620  
 CTCCTTGTTA TGACAAATTG GATTTGGTAT ATGTGTATAA ATAGGGTTTG CACCACAATC 10680  
 ATTCAATTTA CTCATCAAAT TCGCTTTAGC CGCAATGGTC GGCGTTTGAT ATAAATCTTC 10740  
 30 TACACAGTCA ACAATACTG CGTTATTTCG GTATTCTTTT TTCCAAGTCA TAAGACGATG 10800  
 CGCTACAAGT TGCCATAACA CAACTTCATT TCTAGTCGCT TTACCAATAG TTGATACTAA 10860  
 35 ATGTCCTAAG TGATTTACTA CAACGTAATA TTTAAGACGA TGCCATGCTT CATCATGTGC 10920  
 ATATACAACA GGGCTTGATG CTGCCACAAC ATTTGGCACA AGCTGTTTTT CAGTAGCAAT 10980  
 CGTTCTAGAT AGACAAATGC CTTCAAGATC TCTGACAAAG CATACGTCGG GTATGCCATC 11040  
 40 TTTTAATTCA ATTAATGTAT TTTGTACATG TGCTTCTAGA CTAATGCCTG TGTTACTAAA 11100  
 CAGCTTTAAT ATCGGCAATA ATGTACGATT CAAATAACAT TCAAGCCATG CTTCTGGTGC 11160  
 TAAACCACTT TGCTCAATCA CTTGTGATAA CTTAGACATC GGTGAATCAG GCATCGTTTT 11220  
 45 AAATAATGAC GCCAATACAT GAATATCTTT ATCAGCATGG TAATTCGGTA TCCCTTCACG 11280  
 AACAAATCATG GCACTATTTG TTAATAAATC CATTTTCAGG TCAACTGTTT GCCCTAATGG 11340  
 50 ATTCGGTAAC AATGCACGAT ATCCTTCTTC AAACATCAAT TTAATAATGG GTGTTTCAAC 11400  
 CTCATCTTTG ACTGATGCGA TAACTGCGC GGCATCAATT GTCCGTTCAA TCTGTTCAAG 11460  
 55 GTCATTCGTA CGTATAAAAT TAGTGATTTT AACGTGTATC GGTAATTTTA AATAAATGTT 11520

	GCCAAAGGTCT TTTATTAAAC CTTGTTCACT ATATTGCATA TACTGTGGAT GCTGTCGCAA	11640
	CACATTGATT TGATAAGGAT GTGTTGGTAA TAAAAATAAA TCTTTGGGTA TCTCTGATAT	11700
5	ATCTATGTCT GCTAATTGAT ACAACACTTT CTCAACCTGA TCTTCTTTAC CTTCTACATA	11760
	GCGCGTGAGC AGAACATCTT GATGCACAGC TAAATAATGC AATTGGAATG ATGTATGACA	11820
10	TTCCGGTGCA TATTTCTCTA AATCTGCTTC TGAAAACCCA CTTGCACTCT TAGGAGTCGG	11880
	ATGAAATGGA TGACCTAAGT ATAAAGATTG TTCTGAAACG ATATAACGAT CCTCTACGTA	11940
	GTCTATTGTG TTACTTTGCA AATAACGTGC CGTGCGATGA ATGCTATTAT CGATGTCAGA	12000
15	CATAAATTTGC GCCATATGTT GTTGCACTGC CGTTTGATTA TCTGCACTTT GAGCCATATG	12060
	TTGCAAAATA CGCGCAATTG CTTCTTTATA AGTTGTTATT TTTTACTTT TTCCATCGAT	12120
	AAGCCATACC TCTGGATGAT ACATATGATG CCCCATCGCA GACCAATAGC GAAATTCACC	12180
20	CGTTAAAGTT TCGAGCTCTG ATAATTGTAT AGACCATTGA TGATTTTGAG GTGGTACTTG	12240
	ATATAAATTT TCTTCTCTAA AATATTCATT TAAATGCGT TCGATAGCCG CATACGCTGC	12300
	ATGTTGTATT AATCTTTAT TTTGCACTTT TTTGTTTCAA CTCCATAAT TTCATTAATG	12360
25	TGTGATCGTT GATTTGATTA GTGATGGTTG AACAAATTAA AAATAAACTA CTTACTGCAA	12420
	ATACTACGCC CATAACGATA AACGTAGTAG CTGGTGTAGT ATAACCTGTA ATGGCAGCGC	12480
30	cACTaAGACT GCCAATAATT TGACCAACAA CTAACATACT GTTCGTCGTT CCAACAAATG	12540
	TGCCTTTAAG TTGTTGATGA CACGCATTCA CGACAACAA CATGACACTT TGAATCAATG	12600
	CACTATATGT TAATCCTTGA AGTATTCTTG CAGCCATTAA AAACCTCTATA TTCGTCGCTA	12660
35	AACCTTGCGAG TATCGCACTA CAACCACATG CAATCGTGCG AAATATATAT ACTGATTTAA	12720
	CATATGATTT ATCATTAAAG CGTCCCCATA AAGGCGCGCT TAATATCGAA GCCGTCCAAA	12780
	ATGCGGACTG TAAAAATCCA ATCACACTAC GGTCATCTAT CGCTGTATGA TTCACTGATG	12840
40	AAGCAAGTGG TGATAATGCA GTTAGCATGC CATACATAGC AAAGTTTGCT AAAACGCCAA	12900
	CGATAATAAA TCGACATGTT TGTGTGTGC ATAATAGACA TTGAAATGAA CGGCGAATAC	12960
45	CTTTATTAAT ATTTGGTGTT TGTGATTTTG GCATATGTGT CGTTTCAATC AATTTTAATG	13020
	CACCGAAAAT ACAGACAATA AAAGTAATAA CGGCAATACT CATCAGTAAC GCACTAAAC	13080
	CTAATATCGA AGCTGTAACA CCGCCAATTA ATGGCCCCAC AAGAGACCCT GCGCTGACTG	13140
50	AACTTTGCAG TCTTCCTAAT ACCTTTCCAC GATCTTCAGC TGGCGCCTCT GCACTCGCAA	13200
	ACGCACTTGA TGCATCAACA ACACCACCAA ATAGTCCCTG CAATAACCTC ACAAGTACAA	13260
55	ACTGTAATGG TGTCGTACAC AATGCCATTA AAAATAAGCA TACCGCCAAA CCAAGTAACG	13320



	ctATCATCGT CGTTACAGCT GGAGCAGCAA TCGCTATACC ACTCCACAAC TGTATTTCTA	13440
	CGACTGATAG ATTTTGTAGT GATGCCATAT AAATTGGCAA TAATGGCACA AGTACTGTCA	13500
5	GTCCAGCAAT CGCTATAAAC TGA CTGAGCC ATAAAATGCG AAAGTTACTG CGCCATATAG	13560
	ACTGATTAAT CATATGTCAC CATTGGATTT GGTACGGTAG TTAAACCTGA AGGCATACTA	13620
10	CCTCCACCAC TATCACGTTG ATATAGCAAT GGTAATAAAA TTTGTTTGAA TGGCCACGTC	13680
	TGTTTATCAA ATAAAATGTG TCTGACAGCT AGCTGATCAG TTGTAACCCA GGAAATAGTT	13740
	GCCACTTCAT TTTTTAAAAT TTGTTTTAAC AACGACATAA GTTCATGCTC ACTTACACCA	13800
15	AATAAATCTT GAATTGCATC AATAATGGCA TATAGATTTA CCGATACAGC TAATGTTTGA	13860
	AAATAAGCAA AGAATGTTTC CAAATCCTCA TTAATTAGCG TATTAGGTGT ATCTTCTCTG	13920
	ACGACATACT TCGGCAATGA AAGCTGATGT GCTGTTAGCC ATGGTTTATA AATTCTGACA	13980
20	GTATCATGAT CACGTAACAC GCATTTTTGT ACACGTCCAT CTTCAAATGA CAACAATATA	14040
	TTTTGACCAT GCAACTCTGG TAATGCGCCG TATTGCATAA ATGATAGTGT TACCTTTAAA	14100
25	AAGACTTGCG CGATATCTTC AAATAACGTC ATGACATCAT TTTTAGAAAT ATTATCTTTT	14160
	CCACAAATCA TTTGATATAA AGTGCGATCA TTTGCCGCGA GTGCTGCCAT TGACACTAGC	14220
	TGTTGCGTAT CATTTTTGGC TAGCACTTCG GGATACTTTC TTAGCTGAAC AGTTAGATGA	14280
30	CCTAATTGAT CTTTGAAAAT ATCATTATCT TGACCCATAT ATGACCACCA AGCTGTTTCA	14340
	TCACAAACCA TGACATACTT AGCTAGTGCT TCATCTTTTT CTATAAGCTG ACGTAATAAT	14400
	TGTTCTGCTT GTTCTCCGTT TTTCATGTAA CGCGTAGGCG TTAGCCTTAA TGCGCCTAAT	14460
35	GACTGCATTG CAAATGGTAC TTTGACATGG TTATACGGTG CGCCAATATC AATTAATGAA	14520
	CGCATACTTG AAGACGACAG ATAATCTCCA AATTTTAAAC GTAATAGTAC AACCAACTTT	14580
40	TCACTAATCT CTTTCGCAAA GACGTTCCGC AGAATATGCT GATATTGCCA AGGATGTACC	14640
	GGAAATAGTA CATAGTCATC TATTGATAAC CCTTGATCAT TTAACATGTC TGTCGCTTGT	14700
	TCTTTTATAG GTACTGTCAA ATTTTCTAAT TCATCGATAT TTGCAGTATC GCCATGAATC	14760
45	ATATGTGTCT TTTTAACTGC TGCAACCATT AAAGGAAATG ATTGATTTAA TTCAGCTTGA	14820
	TACACTTGAT AATCCGCTTC TCTTAATCCT CTTTTTCTT TAGCTAATGG ATGAAATGGA	14880
	CGATCTTTTA AACTTGCAAA CTGCTCTGAC ATCACAAAAG GATGTGACGC TAAATCTAAT	14940
50	TCTGATAATT GTTTAGCAAG CTGTGTGGCA GCAAGTAGTCA GTCCTTCTTC AACGCGAGCC	15000
	ACTTCCCATC CATGACTTAG ATCACAATTC ATATTAGCAA TTGTTTGCCA AAATTCAGCT	15060
55	GCCGTTAAAG GTTGCTTAGA CACCCTTCCC TCTATCGTAA TTGGTTGTGA ACTTTCGTAA	15120

	TATATCAAAA GCGTTTGTCC GTTTTCTTTA GTAATCTCAC TATTCGATAC AATTCGGGCT	15240
	ATATCTTCAA ATAATAATGC ATCAACTAAA TCTCTTAATA TTATCGCTTG TGCTGTATTG	15300
5	ACTGCTGTAT GATTCTGCAA TGTTCAGACA CCTCGCATTG TTAATATAGG TTCAATGTTG	15360
	TCCCAATATT TTGTTGTTGT GCCTGTTGAT AAATAAAATA AGCACTTGAA ATATCTTCGA	15420
10	TAGCCATACC CATCGGATTA AGTAATATGA TCTCATCATC GTCCTCACGT CCTGGTATGT	15480
	CACCTGTCAC AAGTTGTCCT AGTTCAGCAT GAAGAGCTTC TTTGCTGAAT TTACCTTCTA	15540
	ACACCAATTG GTTAATAGTT TTCTTTTCTC GATTACATTG TGACCAGTCA TCTACTACGA	15600
15	CTTTGTCAGC TTTAATAAAG ACTTCTTTAT GCACATCCAT GATAGAAATG TTGCTAATAA	15660
	ATGCACCCTT TTGTAACCAA TCATATTCAA TGTATGGTTG ATCCGTTACG GTACATGTAA	15720
	TGACTACTTC ACCATTTGAT ACTGCTTCTT TAGCATTTTC TGTCGCAATA AAATTAATTT	15780
20	CCGGACGCTG TTGTTGCCAT CTATCAACAA AGCGTGCACA TGCTTCAGAG AATTGATCGT	15840
	AAACAAACAC GCGTTCAATA TGATCGAATT GCTCTAACAT ACTTTGTAAT TGCTTGTCTC	15900
25	CGATTAGCCC GCATCCAATG ATTGTTAAGT CTTTAAATCC TTTTITAGCC AAATGCTTTG	15960
	CTGCAATCAC TGAAACTGCT GCAGTACGCA TACTACTAAT TAAACTTGCT TCCATAACTG	16020
	CAATTGGATA ATTCGTTTCT GGATCATTCA AAATAATGAC GCCACTTGCA CGCTCCATAT	16080
30	TACGTTTCGA TGGATTGTCG TGCTTACTAC CTATCCACTT AATACCTGAA ATTGCGTGTT	16140
	CACCACCGAT ATGACTTGGC ATTGCAATAA TTCGATCTGC GATGTGTCCA TTTTCAGGAT	16200
	CCTGTCTTAA ATACGGCTTA AGCGGTTGTA CAAAATCATT GTGCGCATGG GCTGTTAATG	16260
35	CTTCTGTTAA TGGCTCCACA TAACTTGTG AATGATTACC TCCCGCTTGT TCAATATCTG	16320
	ATCTATTTAA ATACAACATC TCTCTatTca TTCTGaTTTA ACTCCTTGTC TTGATTTTAT	16380
	TTTTTCTAAC CATGTATCTG AATAAACTAA ATCTAAGTAA CGATCGCCTC GATCTGGTAA	16440
40	AATCGTGACA ATTGTTGCAC CTTCTTCAAT TGACGTTATC AACTGCTCAA TCGCTGCAAT	16500
	AATCGAACCT GTTGAAcCTC CGGCAAATAT GCCTTCATAA TCAATCAGTT TTCGACAGCC	16560
45	CAAAGCAGAT TGATAATCAT CTACATGGAT CACTTGATTA ATTTCTGATC TATTCAATAT	16620
	TTCCGGTACA CGACTAGCAC CGATACCAGG TAATTCTCTA TTAATAGGTT TGTCAACAAA	16680
	AATGACTGAC CCTTTCGCAT CAACAGCAAC AATTGTGCG TTTGGATGCA CTTCTTTTAT	16740
50	TTTTCTACTC ATACCCATAA TGCTACCTGT CGTGCTGACT GGCGCGACAA AATAATCTAT	16800
	AGGTTGCTTA ATTGTTTCAA CAATCTCTGT GCCTGCACCA TGATAATGGG ATTGCCAATT	16860
55	TAACTCATTC GCATATTGAT TAATCCAATA TGCATCGTCA ATAGTGGCTA ACAGTTCCTG	16920

TACATTGGCA CCATAACTTT TAATAATTTT CAAATTTGTT GGTGATATTT TAGGATCAAC 17040  
 AACACACGTG AGTTTTAATC CCTTGATTTT AGCTATCATT GCCAACGCAA TGCCTAAATT 17100  
 5 ACCAGAAGTA CTTTCAATTA AATGTGTATT CTCAGTGATT AAACCATGTT TAATACCATG 17160  
 TTCAATGATG TACTTGGCAG GTCGATCTTT CATGCTGCCT CCAGGATTCA TATACTCTAA 17220  
 10 CTTTGCAAAC ACTTCATGTT TCGGAAATAG TTGATGAAGT TGAACCATAG GTGTTTGCCC 17280  
 TACAGAATCT AACAAATGAAT CGTGACATG 17310

## (2) INFORMATION FOR SEQ ID NO: 24:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5423 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

ATACTAGTAA GCGCATCGGT TATTGACATC GAATTCAACT TTAACAGTTT TCATGTTCCG 60  
 25 TGATGTTTca ATAGAATGTG TGTGTTGTAC TTGCGCATTT ATATTTCCAC CTAAATTACT 120  
 TAAGTTTCCT GTAATACTAG AAATGTCAGG TCGGTTTAAT GTAGGTTGAA ATGCATCAAC 180  
 30 TACTTTATCT GCAACATTAG AACATTACG GATAACTTTA CTTGAATGAT TATCTATACC 240  
 TTTAACGAAA CCTAACATTG AATACATACC AACATCCATG AATTCACGTG AAGGTGAGTG 300  
 AATACCTAGC GCTCTTTTGG CTGCATTTAA AGCACCTTTT GCTACACTAG CTGCTTTTTC 360  
 35 AGCTAAGTCT CTAGCCATAT TACCAATACC TCTCATCAAA CCACGGATCA TATCAGCACC 420  
 TGCTGATACA AAGTCATCCA CAAAGCTTTT AACTTTATTT ACTGCATTTG TCATACCTTG 480  
 ACTAACTTTG TTTACAACAT TAACGAATCC TTGAACAACT CTATTAACAA rGTTAATTAG 540  
 40 CGTACTtTGtT ATAGTAGATA CCCaTnGCAT ACCTTTAGTG ACmATGAAGT TCCAAGCTTG 600  
 AGACATTTTG TCTGATATAG TTGAAACAAC TTGTGTGAAT ATGCTTACAA CTTTATTCCA 660  
 AATTGTCGTT AATATACCAG ATAAGAAACT CCAAATCGTA TTCCATATAT TAGAAATAAA 720  
 45 ACTCCATGCC GCTTGTAACG CAGTAGATAT AGCTGTAGTG ATAGCGTTCC AAACCTTAGT 780  
 TGCCACAGTA ACTATAGTGT TCCACAACGT TTGTAAGAAC GTCCAAATAG CGTTCCAAAT 840  
 50 TGTATTGCG ATAGTCATAA TTGTGGTAAA CACTGTAGTT ATTACAGTGA CTAACAAATT 900  
 CCAAATCGTA GTAGCGATTG TAATTATCGT ATTCCAGATT GTACTTAAGA ACGTCCAAAT 960  
 AGCTGTCCAT ATCGTCATAA CTATTGTCAT TATCGTCGTG AAAACAGTTG TAATGATTGT 1020

	ATAAGCGACT ATTTGATTCC AAACAATCAT TATAAAATTG TAAACATTTCG ATACTGCTGT	1140
	AGTGATAGCT GTTAAAATAG CATTCCATAC AACCGAAGCT ACAGCTTTTA ATACATTCCA	1200
5	AACATTAACC ATAAACGTTT TTATCGCATT CCAAGCATTT ATAATAAAGT TTCTGAATCC	1260
	TTCATTTTTA TTCCACAATA AAACGAATAT AGCTATTAAT GCAGCAATTA CACCAATTAC	1320
10	TATTGTTATT GGACCGCCTA AAATACCAAA CACAGTACT AGTCCTGTGA TAGCATTCT	1380
	AATTAATCCA ATCTTACCGA ATAACAATTG GAATATAACT GATATAATTT TTAATGGTCC	1440
	TTTTAATAAC ATGAACGCAC CTTTTAAAT TGTTAATCCC GCTCTTAATA AACCGAACTT	1500
15	ACTTACTAAT GCAATGrTTC TACCTATTAA TCCGCCACCC ATAAAGTTAG ATACAGCAAG	1560
	AATAATCGGT ATTAATAATC TAAATGCACC AACTAAAGTT ATAATGACAC CAACTAATTG	1620
	TGCTGTAGCT GGATGCGCCT CAAACAAGTT AGCTATCCAA CCAGTTATTG CAACTGCAAC	1680
20	GCGTAATACT GCACTAGCTA TAGGAGCCAT TGCTGTTGCG AATGCarmTA ATCCTCTTGC	1740
	GATGTTTCCA ATCAATTGCA TTATTAGTGG TCCATTTGTT TGTATATAAC TGACAAAGTC	1800
25	TTTAAACCCT TGAGATTGTC CTAATTGTTT AGACCATTCC CTAAACTTAG CTGTCATTG	1860
	TTCAAGAGAT TGGAATATGC CAGTTGATGA TCCGCTGAAT GCATTCATCA AATTGTTAAT	1920
	TCCAACGAAA ACATTTTTGA AAATATTACC AATGATAGGT AAGTTTGTTC TTGTGTATTC	1980
30	AATAAACGA GTTATCGAAT TTTCTCCAGC TGCACTATTA GCCCAGTTAG AGAAAGATTG	2040
	ACCTAATCTA TCCAACCAAT CAGCCGACCA TTGAAACAGT GGTGCTAATT GCGTGAATAC	2100
	ATTGACTAAT CCGTCACCAA AACCACCTGC AGCACTTAAT AGCTTGTTAA ATACCGAAAC	2160
35	ACCCGTTGTA TTCATCATAT TAAAGAATCT TGAAGCTACA CTGCTATTTT CAGCCCATT	2220
	AAGCAGCCTT TGAGACGCTT CTTCCATTCC TCTTGAAATA CCACTAAAAA ACGGTTGTAA	2280
	GCTCTGCATT GCAGTTTTAA CAGTATTTAA ACCATTGCA AGAGTTGTGA AGATAGCGGA	2340
40	TTGATTTTGC TTTATAATAT CAGTCCATGC TGACTTTACG CCATCTAACG CTTTTTTGTA	2400
	TTCGTTTGTT GCTGAGCTAG CTTGTAAAGT GCCATCATTAGCATCTTTA TAGCGCTGAT	2460
45	AGCCATTGCG CCAAACGCTA CAAATCCTGC TCCCGCTATT GCTACGGCAC CACCTAAAGC	2520
	AAGTACACCA CCAGTTAACA CTTTGATAGC GTTTAATAGC GCAAATACTA CAGGTACTAC	2580
	GCTCGCTATT ACAGGTATTA AGATACTAAA AGATGATGTA AGTAATCCAC CAACCATATT	2640
50	AGAACCTACA GTACCGAACA CACGGAACAT ATTAGCTAAA TTCCCCATCT GTCTTTGAAA	2700
	ATTGTCATTT GCTTTTATTA TGTAGGCATA AGCTTTCTTT AAACCATTAG TATCGACATC	2760
55	TACCTTTGTT GTTTTTTTGT TCGGCAATGC GTCTAATGAT TTTTAAACG CATAAATAGT	2820

	AAGTTCTTCT	TTAGTACGTT	TGATTTTAGA	GTTAGCAACA	CCATTGTCCA	CGTCTATAAT	2940
	AGCTTTGGCT	TTAGACCTAT	TTAATGCTTC	GAGACTAGCT	TTAGATACTT	TTAACACTCG	3000
5	ATTGAATTTA	CTGTTATCTG	CATTGACGTC	AATATTGACA	CGTTTCTTTT	CTAATTCTGA	3060
	TAATTTAGCT	TCTGTTTCAG	CGATATCTTT	AATCAACTTT	TGTTTTTGCA	ACTTAACTTC	3120
	TGGTGTAAGT	TCTTTAGAGT	TTAGTTTGTC	TAGTTCAAAA	TTGATTCTA	GTACCTTTTG	3180
10	TTGTAAATCT	TGTATACTAG	CATCTAATTT	AGCTTTTACA	TTTTTGTTAC	TAAAGGCATC	3240
	TAAAGACTTT	TTAGCAACTT	TGATAGTTTT	TTGTAAATTT	TTATCGTTAG	CGTTTAATTC	3300
15	AACATCTTTA	GTTTGATCTG	CTACTCGTTT	AAATCTTTGC	ACAGACTTAA	CCGCACTATC	3360
	AATTTGCCTT	TTGAATTTGG	CTACACTAGC	TTCAATAGTC	GCTTTAATTT	TATATTCGGT	3420
	CACATTAACA	CCTCTCTTTC	TATTGCTTAT	TAAATTCTGC	TATAACTTTA	AAGAATTCAT	3480
20	TATTTTGTGG	TTGCTATTCA	TCACGTTGCG	TACTAAATCT	TATATCTTTA	CCTTCGTTAA	3540
	GCCGTTGGAT	ATTTTCTTCA	TAAGGCAATA	CGTCGTTTGC	ATTGTTAAAA	ACATATTCCT	3600
	CTTTAGGTTT	ATTTTCTGTC	CCAACATTTT	TAGTAGCTGC	AGCATCACGA	ATAGCAAACG	3660
25	CAAGTTTGTA	ACGTTGCAAT	TCTTGGGTTA	GCATTTTATA	CTCTTTTCGA	TACATTGAT	3720
	AGTTATATTC	TGTTAATGTC	ATTTGCTCAA	TAACGTTCAA	ATCTGTAATA	CCAAGTGTTG	3780
30	ACATACAAGT	TATAACGATT	CTGTCGTAAG	TTATTAGGCT	TCGCTGGTT	TTTCTTCCGT	3840
	TTCCACTACT	TCGACTAGGT	TTGGGTCAT	AGGTCGCTTT	CCCAAATCCG	TTAAAATATC	3900
	CGAACCGAAT	TCTTCTAGTC	CGATATTTTC	TGCGATTTCA	TCTAATGCTT	CATCAATGTT	3960
35	ATTAATAGTA	ATTGCTTGTT	TTTTTAAGTG	AGATGTAGCT	GCGATTAAAA	CTTCGCCAAT	4020
	CACAACCGGA	TTCCACTTTT	CTAAACCTAC	AGGCAACATT	GATACACCTT	GACCGATAGA	4080
	AGCTTGTTCA	ACTTTTAAAC	CTAATCGGTT	ATCGATTTCT	CTTAAAAATT	TAAAACCAAA	4140
40	ACTTAATTCT	AATGACTTTC	CGTTAATTTT	TACATTCTTA	ACTTAAAATC	TCCATTCTTA	4200
	ATTAATTTAA	ACAAAATAAA	TAAGCTTAAC	GCCCTATTTT	TATACCTCTC	TTGGTGCAAC	4260
45	CGGTGGTGAA	TCTACTTTAG	GTTGTGGAAT	TGCTGTTAAA	TCTTCGCCAG	TTAATGCATC	4320
	TGCTTTTGTA	GTGTCGTGGA	ATCTGTATCT	AGTCGCCTTA	AGTTTCTTTG	TTACAGCCTC	4380
	AGGTAGTGTT	GCAAATCCAC	GTTGGAAACG	ACCATTCACT	CCATATTCAT	ATTCATATTC	4440
50	ATCAATACCG	TTAGCTTCTG	CTTTTAATTC	AAATTTATTG	TGGAAACCTT	GGAAATATTT	4500
	CGCTTTTAAAT	TTAGCGGAAT	CCCCATTTTT	GCCTGGTATT	CTACTTTCAA	CTTCCCAAGC	4560
55	TTCATACAAT	ACGCGATCTA	CAACTGCATC	TTCAATTTCA	TCTGCAAAAT	CGTCACCATA	4620

GTCCATTGTA TCCTCTGTAT CTGTATCAGC TTCATGTGAT AAGCCGTATT CAGTTAAAAA 4740  
 AAGCATTTTA GTAGCATCTA CTTTTTCGCC AGCTTTTCTA AATAAAATAA TACGATCATT 4800  
 5 ACTATTTTTC ATATTGCCA TTCAATATTC CTCGGTTTTT TAAATGTTT TGTAAGATAT 4860  
 CGTTACTGAT GTGTGTAGCA ATTCTTGATT GGTAGTATCA TCAACTAACT GTGTGATGTT 4920  
 10 AGTATCTTCT TCTTCAAAGT CATAATCGTT TGTTTTAACG CTAGGTGTTA AATCATCAAT 4980  
 ACATCTTTTA ACAAGTCCGT CATGATGTCC TAAATCATCG CTTACACTCC AAATATCAAT 5040  
 AACTAAATTC GTATCGCCAG AATAACTATC AAACGTGTAC TTA CTCTTCTAT TTGACTCCGG 5100  
 15 CATTTTATT ACAAAAAAG GATACGGAAT CTCTTGTTGC ATCTCTTTAC GAGAAATAAC 5160  
 AGGGAATCCA TATCCTTGTA GCGTTTCATA CGCTTTATTA TAAAGTTGTA AGTTCGGTGT 5220  
 CATGCTTTTA TCTCCTATTC AAACAACGCT TTCAATTCTT CTACAGTTGA TTTCTAATC 5280  
 20 ACTTCGTATA CCGGCCACAT AAAAGGTTCA GCCTCCATGT ATCGAGTACC AAATTCTAAG 5340  
 AAACCACTAT AAGCTGCGTG CGATGTGATA GTGTATTGCA AATCGCCAGT TTTTATAT 5400  
 CTGATATTGC GTGATAAATT ACC 5423

(2) INFORMATION FOR SEQ ID NO: 25:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 6251 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

AAACGCAGAT GTTCAATTAG AACCACTCTA TCGTATTAAG GAAGGTATTA AACAAAAGCA 60  
 AATAACGAGAC CAAATTAGAC AAGCGTTAAA TGATGTGACA ATTCATGAAT GGTAACTGA 120  
 40 TGAAC TAAGA GAAAAATATA AATTAGAGAC CTTGGACTTT ACTTTGAACA CATTACATCA 180  
 TCCTAAAAGT AAAGAGGATT TATTACGTGC TCGTAGAACC TATGCATTTA CTGAACTGTT 240  
 45 TTTATTGAA TTACGTATGC AATGGCTAAA TAGATTAGAA AAGTCATCTG ACGAAGCAAT 300  
 TGAAATTGAT TATGACATAG ACCAAGTTAA ATCATTTATT GATCGTTTAC CTTTGAAC 360  
 AACTGAAGCA CAGAAATCCA GTGTTAATGA AATTTTAGA GATTTAAAAG CACCAATACG 420  
 50 TATGCATCGA TTA CTCAAG GTGATGTAGG TTCAGGAAAA ACAGTAGTTG CTGCAATTTG 480  
 TATGTATGCG TAAAAA CTG GTTATCA ATCAGCATTG ATGGTACCAA CTGAAATTTT 540  
 AGCAGAGCAA CATGCTGAAA GTTTAATGGC TTTATTTGGA GATTCTATGA ACGTTGCATT 600

	TACGATTGAT TGTTTAATTG GAACCCATGC TTTGATTCAA GATGATGTGA TTTTCCATAA	720
	TGTTGGTTTA GTAATTACAG ATGAACAACA TCGATTGGT GTGAATCAAC GCCAGCTTTT	780
5	AAGAGAAAAA GGTGCAATGA CGAATGTGTT ATTTATGACA GCAACGCCGA TACCAAGAAC	840
	ACTAGCAATA TCAGTTTTTG GTGAGATGGA TGTGTCTTCA ATTAAACAAT TACCAAAAGG	900
10	TCGTAAACCT ATCATTACTA CTTGGGCAAA GCATGAGCAA TACGATAAAG TTTTGATGCA	960
	AATGACCTCA GAGTTGAAAA AAGGTCGTCA AGCATATGTC ATTTGCCCGC TAATAGAAAAG	1020
	TTCTGAGCAT CTCGAAGATG TTCAAAATGT TGTGCGATTG TACGAGTCTT TACAACAGTA	1080
15	TTATGGTGTT TCCCGTGTAG GGTATTGCA TGGTAAGTTA TCTGCCGATG AAAAAGATGA	1140
	GGTCATGCAA AAGTTTAGTA ATCATGAGAT AAATGTTTTA GTTTCTACTA CTGTTGTGTA	1200
	AGTAGGTGTT AATGTACCGA ATGCAACTTT TATGATGATT TATGATGCGG ATCGCTTTGG	1260
20	ATTATCAACT TTACATCAGT TACGCGGTGCG TGTAGGTAGA AGTGACCAGC AAAGTTACTG	1320
	TGTTTTAATT GCATCCCCTA AAACAGAAAC AGGAATTGAA AGAATGACAA TTATGACACA	1380
	AACAACGGAT GGATTGAAT TGAGTGAACG AGACTTAGAA ATGCGTGGTC CTGGAGATTT	1440
25	CTTTGGTGTT AAACAAAGTG GaTTGCCAGA TTTCTTAGTT GCCAATTAG TTGAAGATTA	1500
	TCGTATGTTA GAAGTTGCTC GTGATGAAGC AGCTGAACCTT ATTCAATCTG GCGTATTCTT	1560
30	TGAAAATACG TATCAACATT TACGTCATTT TGTGAAGAA AATTTATTAC ATCGTAGTTT	1620
	TGACTAATTG CCATGCTGAT TTGTCAATTT GAGTGCAACa CTTGTTAAT TGAGTGATAT	1680
	GACACTTGAA CTATTTAAAT GTAAAGTGGT ATTTTAACAA TTTATAAATT TTCGACTAAA	1740
35	TAATAGCTAA ATATTACAGT TATTTGTTGA GTCGGTTAAA TAGAAAGTGT TATGATATGT	1800
	GAGGAATGTT TAAGACTAGG TACTAAAAAA TGAGGGGTGA GACGTTGAAA CTAAAGAAAG	1860
	ATAAACGTAAG AGAAGCAATC AGACAACAAA TTGATAGCAA TCCCTTCATC ACAGACCATG	1920
40	AACTAAGCGA CTTATTTCOA GTGAGTATAC AAACAATTCG TTtAGaTCGC ACTTATTTAA	1980
	ACATACCAGA ATTAAGGAAG CGTATTAAAT TAGTTGCTGA AAAGAATTAT GACCAAATAA	2040
45	GTTCTATTGA AGAACAAGAA TTTATTGGTG ATTTGATTCA AGTCAATCCa AATGTTAAAG	2100
	CGCAATCAAT TTTAGATATT ACATCGGATT CTGTTTTTCA TAAAACTGGA ATTGCGCGTG	2160
	GTCATGTGCT GTTTGCTCAG GCAAATTCGT TATGTGTTGC GCTAATTAAG CAACCAACAG	2220
50	TTTTAACTCA TGAGAGTAGC ATTCAATTTA TTGAAAAAGT AAAATTAAAT GATACGGTAA	2280
	GAGCAGAAGC ACGAGTTGTA AATCAAACCTG CAAAACATTA TTACGTCGAA GTAAAGTCAT	2340
	ATGTTAAACA TACATTAGTT TTCAAAGGAA ATTTTAAAT GTTTTATGAT AAGCGAGGAT	2400
55		

	TTAGAAGCCG TACAAAAGGC TGTGAAGAC TTAAAGATC TAGAAATTAT ACTTTTCGGT	2520
	GACGAAAAAA AGTATAATCT GAACCATGAA CGAATCGAAT TTAGACATTG TTCTGAAAAG	2580
5	ATTGAAATGG AAGATGAGCC TGTAGAGCG ATTAAACGTA AAAAGATAG CTCAATGGTA	2640
	AAAATGGCTG AAGCTGTGAA ATCTGGTGAA GCAGATGGAT GTGTGTCAGC AGGTAATACT	2700
10	GGTGCTTTAA TGTCAGCTGG TTTATTCATT GTTGGACGTA TTAAAGGTGT AGCTAGACCG	2760
	GCTTTIAGTAG TAACATTGCC AACGATTGAT GGAAAAGGTT TTGTCTTTT AGACGTTGGT	2820
	GCAAATGCTG ATGCTAAACC TGAACACTTA TTACAGTATG CGCAACTAGG GGATATTTAT	2880
15	GCTCAAAAAA TTAGAGGTAT TGATAATCCG AAAATCTCAT TATTAAATAT AGGAACCGAG	2940
	CCAGCTAAAG GTAATAGTTT AACGAAAAAA TCATATGAGT TATTAAATCA TGATCATTCA	3000
	TTGAATTTTG TTGGGAATAT TGAAGCGAAG ACATTAATGG ATGGCGATAC AGATGTTGTA	3060
20	GTTACCGATG GCTATACTGG GAACATGGTC CTTAAAAATT TAGAAGGTAC TGCAAAATCA	3120
	ATCGGTAAAA TGTTAAAAGA TACGATTATG AGTAGTACTA AAAATAAATT AGCAGGTGCA	3180
	ATATTGAAGA AAGATTTAGC TGAATTCGCT AAAAGATGG ATTACTCAGA ATACGGTGGT	3240
25	TCCGIATTAT TAGGATTGGA AGGTACTGTA GTTAAAGCAC ACGGTAGTTC AAATGCTAAA	3300
	GCTTTTTATT CTGCAATTAG ACAAGCGAAA ATCGCAGGAG AACAAAATAT TGTACAAACA	3360
30	ATGAAAGAGA CTGTAGGTGA AtCAAATGaG TaAAACAGCA ATTATTTTTC CGGGACAAGG	3420
	TGCCCAAAAA GTTGGTATGG CGCAAGATTT GTTTAACAAC AATGATCAAG CAACTGAAAT	3480
	TTTAACTTCA GCAGCGAACA CATTAGACTT TGATATTTTA GAGACAATGT TTAGTATGA	3540
35	AGAAGGTAAA TTGGGTGAAA CTGAAAACAC ACAACCAGCT TTaTTGaCGC aTAGTTCGGC	3600
	ATTATTAGCA GCGCTAAAAA ATTTGAATCC TGATTTTACT ATGGGGCATA GTTTAGGTGA	3660
	ATATTCAAGT TTAGTTGCAG CTGACGTATT ATCATTTGAA GATGCAGTTA AAATTGTTAG	3720
40	AAAACGTGGT CAATTAATGG CGCAAGCATT TCCTACTGGT GTAGGAAGCA TGGCTGCAGT	3780
	ATTGGGATTA GATTTTGATA AAGTCGATGA AATTTGTAAG TCATTATCAT CTGATGACAA	3840
45	AATAATTGAA CCAGCAAACA TTAATTGCCC AGGTCAAATT GTTGTTTCAG GTCACAAAGC	3900
	TTTAATTGAT GAGCTAGTAG AAAAAGGTAA ATCATTAGGT GCAAAACGTG TCATGCCTTT	3960
	AGCAGTATCT GGACCATTC ATTATCGCT AATGAAAGTG ATTGAAGAAG ATTTTTCAG	4020
50	TTACATTAAT CAATTTGAAT GCGGTGATGC TAAGTTTCCT GTAGTTCAAA ATGTAAATGC	4080
	GCAAGGTGAA ACTGACAAAG AAGTAATTAA ATCTAATATG GTCAAGCAAT TATATTCAAC	4140
	AGTACAATTC ATTAAC TCA CAGAATGGCT AATAGACCAA GGTGTTGATC ATTTTATTGA	4200
55		



	AACATCAATT CAAACTTTAG AAGATGTGAA AGGATGGAAT GAAAATGACT AAGAGTGCTT	4320
	TAGTAACAGG TGCATCAAGA GGAATTGGAC GTAGTATTGC GTTACAATTA GCAGAAGAAG	4380
5	GATATAATGT AGCAGTAAAC TATGCAGGCA GCAAAGAGAA AGCTGAAGCA GTAGTCGAAG	4440
	AAATCAAAGC TAAAGGTGTT GACAGTTTTG CGATTCAAGC AAATGTTGCC GATGCTGATG	4500
10	AAGTTAAAGC AATGATTAAA GAAGTAGTTA GCCAATTGGT TTCTTTAGAT GTTTTAGTAA	4560
	ATAATGCAGG TATTACTCGC GATAATTTAT TAATGCGTAT GAAAGAACAA GAGTGGGATG	4620
	ATGTTATTGA CACAACTTA AAAGGTGTAT TTAAGTGTAT CCAAAAAGCA ACACCACAAA	4680
15	TGTTAAGACA ACGTAGTGGT GCTATCATCA ATTTATCAAG TGTTGTTGGA GCAGTAGGTA	4740
	ATCCGGGACA AGCAAATAT GTTGCAACAA AAGCAGGTGT TATTGGTTTA ACTAAATCTG	4800
	CGGCGCGTGA ATTAGCATCT CGTGGTATCA CTGTAAATGC AGTTGCACCT GGTTTTATTG	4860
20	TTTCTGATAT GACAGATGCT TTAAGTGATG AGCTTAAAGA ACAAATGTTG ACTCAAATTC	4920
	CGTTAGCAGC TTTTGGTCAA GACACAGATA TTGCTAATAC AGTAGCGTTC TTAGCATCAG	4980
	ACAAAGCAAA ATATATTACA GGTCAAACAA TCCATGTAAA TGGTGGAATG TACATGTAAT	5040
25	ATATTTGAGC TAAAGCTCAT TGACGCAGTG GTTGACTGGT CATCCAATGG AGAATTGTCT	5100
	GACCTAGTCA ACTTTGCGGG GGAAATTCTA AGCAACCTAG ATAAGGTTCC AGAATTTCTC	5160
30	CCTAAGAAAC ACTAATCAAT aAATTGWTA GTGTTTCTAA AATTTCTACT TGTTTTTTAG	5220
	AATTTAAAT GGGAAATAT AGTAGTCTAT GTATAGGCAT TTTTAAAGGA GGTGAATCGA	5280
	CGTGGAATAT TTCGATAAAG TAAAAGATAT CATCGTTGAC CgTTTAGGTG TAGACGCTGA	5340
35	TAAAGTAACT GAAGATGCAT CTTTCAAAGA TGATTTAGGC GCTGACTCAC TTGATATCGC	5400
	TGAATTAGTA ATGGAATTAG AAGACGAGTT TGGTACTGAA ATTCCTGATG AAGAnGCTGA	5460
	AAAAATCAAC ACTGTTGGTG ATGCTGTTAA ATTTATTAA ACCTCTGAAA AATAATAAAT	5520
40	CTTACATCTG GGTGTCAGT ATTGTCGACT CAGTTTTTTT CTTTAATTAT CAATAGTTTT	5580
	AACGTAAAT TAAAGATGAT TCAAGAGCAA CACATAAAGG AGATAAAATA ATGTCTAAAC	5640
45	AAAAGAAAAG TGAGATAGTT AATCGTTTTA GAAAGCGCTT TGATACTAAA ATGACAGAGT	5700
	TAGGCTTTAC TTATCAAAAT ATTGATTTAT ACCAACAAGC ATTTTCGCAT TCGAGTTTTA	5760
	TTAATGATTT TAATATGAAT CGTTTAGACC ATAATGAGCG TTTAGAGTTT TTGGGTGATG	5820
50	CGGTATTAGA ATTGACGGTT TCACGATATT TATTTGATAa ACATCCCAAC TTGCCAGAAG	5880
	GGAATTTAAc AAAAATGCGT GCCaCTATTG TATGTGAGCC CtCACTkGTA ATATTTGCGA	5940
55	ATAAAATTGG ATTGAACGAA ATGATTTTAC TTGGTAAAGG TGAAGAGAAA ACAGGGGGAC	6000

ATCAAGGACT AGATATAGTT TGGAAATTTG CTGAGAAAGT CATTTTCCCA CATGTAGAAC 6120  
 AAAATGAGTT ATTAGGCGTG GTAGATTTTA AAACACAATT CCAAGAATAT GTGCACCAGC 6180  
 5 AAAATAAAGG TGATGTAACC TATAATTTAA TAAAAGAAGA GGGACCGGCA CATCATCGTC 6240  
 TATTCAC TTC A 6251

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4920 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 15 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

20 ACCTACTGAA GTTGCTAATT TTTTGGAGCA ACTAAGCACT GAAATTGAAC GTCTTAAAGA 60  
 AGATAAAAAA CAACTTGAAA AAGTAATCGA AGAGAGaGAT ACTAATATTA AGTCTTATCA 120  
 AGACGTGgCA TCAATCTGTA AGTgaTGCTT TGATACAAGC TCAAAAAGCT GGTGAAGAAA 180  
 25 CTAAGCAAGC TGCAGAGAAA CAAGCTGAAG CGATTATAGC TAAGGCAGAA GCGCAAgcTA 240  
 ATcAAATGGT TGGTGACGCG GTAGAAAAAG CACGCCGTTT AGCATTCCAG ACTGAAGATA 300  
 TGAAACGTCA ATCAAAAAGTA TTTAGATCGC GTTTCCGTAT GTTAGTTGAA GCGCAATTAG 360  
 ACTTATTAAA AAACGAAGAT TGGGATTACT TGTGTAATTA TGATTTAGAC GCTGAACAAG 420  
 TGACGCTTGA AAATATTCAT CATTTGCATG AAAATGATTT AAAGCCAGAT GAAGTTGCAG 480  
 35 CAAATGCACA AAATAATGCA TCAAATACAC CAGACAATAA TCAACAATCC AATGATTCAG 540  
 AAACAAC TAA GAAGTAAGAA TTAAATAAAG ACAGACGCGT AATATACATT TAACTTTTCA 600  
 CAGCGAATTA GGTAATGGTG AGAGCCTAGT AAAAGCATGT ATGTTATATC ACTGGCTTTT 660  
 40 TAATATTTAA ATAATGTAAT GAGAGAACTC TAAGTTGAGT TAATAAGGGT GGTACCGCGA 720  
 GCAATCGTCC CTTTTAATTT AACTTAGAGT TTTTAAATTT TTAAAGGAGT GAAAAAATG 780  
 GATTACAAAG AAACGTTATT AATGCCTAAA ACAGATTTC CAATGCGAGG TGGTTTACCA 840  
 45 AACAAGGAAC CGCAAATTC AAAAAATGG GATGCAGAAG ATCAATACCA TAAAGCGTTA 900  
 GAAAAAATA AAGGTAACGA AACATTCATT TTACATGATG GCCCACCATA CGCGAATGGT 960  
 50 AACTTACATA TGGGACATGC CTGAAACAAA ATTTTAAAAG ACTTTATTGT ACGTTATAAA 1020  
 ACTATGCAAG GGTTCTATGC ACCATACGTA CCAGGTTGGG ATACACATGG TTTACCAATT 1080  
 GAACAAGCAT TAACGAAAAA AGGTGTTGAC CGAAAGAAAA TGTCAACAGC TGAATTCGGT 1140

	TTAGGTGTTT	GTGGTGACTT	TAATGATCCA	TATATTACAT	TAAAACCTGA	ATACGAAGCT	1260
	GCACAAATTC	GTATTTTGG	AGAAATGGCA	GATAAAGGTT	TAATTTATAA	AGGTAAAAAG	1320
5	CCAGTTTATT	GGTCTCCTTC	AAGTGAGTCT	TCATTAGCAG	AAGCAGAAAT	TGAATATCAC	1380
	GATAAACGTT	CAGCATCAAT	TTACGTTGCA	TTTGACGTTA	AAGATGACAA	AGGTGTCGTT	1440
10	GATGCAGATG	CTAAATTTAT	TATCTGGACA	ACAACGCCAT	GGACAATTCC	ATCAAATGTT	1500
	GCGATTACCG	TTCATCCTGA	ATTAAAAATAT	GGTCAATACA	ATGTAAATGG	cGAAAAATAT	1560
	ATTATTGCAG	AAGCCTTGTC	TGACGCTGTA	GCAGAAGCAC	TGGaTTGGGA	TAAAGCATCA	1620
15	ATCAAATTAG	AAAAAGAATA	CACAGGTAAA	GAATTAGAGT	ATGTTGTAGC	ACAACATCCA	1680
	TTCTTAGACA	GAGAATCGTT	AGTGATTAAT	GGTGATCATG	TTACTACAGA	TGCTGGTACA	1740
	GgTTGTGTAC	ATACAGCACC	AGGTCACGGG	GAAGATGACT	ATATTGTTGG	TCAAAAAATAT	1800
20	GAATTGCCAG	TAATTAGTCC	AATCGATGAT	AAAGGTGTAT	TTACTGAAGA	AGGCGGCCAA	1860
	TTTGAAGGGA	TGTTCTATGA	TAAAGCTAAT	AAAGCCGTTA	CTGATTTATT	AACAGAAAAA	1920
	GGTGCACTAT	TAAAATTAGA	CTTTATTACA	CATAGCTATC	CACACGACTG	GAGAACAAAA	1980
25	AAACCTGTAA	TCTTCGGTGC	TACACCACAA	TGGTTTGCTT	CAATCAGTAA	AGTAAGACAA	2040
	GATATTTTAG	ATGCAATCGA	AAATACAAAC	TTCAAAGTAA	ATTGGGGTAA	AACACGTATT	2100
30	TACAATATGG	TTCGTGACCG	TGGCGAATGG	GTTATTTCTC	GTCAACGTGT	GTGGGGTGTA	2160
	CCGTTACCAG	TATTTTATGC	TGAAAATGGC	GAAATTATCA	TGACGAAAGA	AACAGTGAAT	2220
	CATGTTGCTG	ATTTATTTGC	AGAACACGGT	TCAAATATTT	GGTTTGAAAG	AGAAGCGAAA	2280
35	GACTTACTAC	CAGAAGGATT	TACACATCCA	GGCAGCCCTA	ACGGTACATT	TACTAAAGAA	2340
	ACAGACATTA	TGGACGTTTG	GTTTGATTCT	GGTTCATCAC	ACCGTGGCGT	GTTGGAAACA	2400
	AGACCGGAAT	TAAGTTTCCC	AGCGGATATG	TATTTAGAAG	GTAGTGACCA	ATATCGTGGT	2460
40	TGGTTCAACT	CTTCTATCAC	AACTTCAGTT	GCTACAAGAG	GAGTATCACC	TTATAAATTC	2520
	TTACTTTCTC	ATGGTTTTGT	TATGGACGGT	GAAGGTAAGA	AAATGAGTAA	ATCTTTAGGT	2580
45	AATGTGATTG	TACCTGACCA	AGTGGTTAAA	CAAAAAGGTG	CTGATATTGC	GAGACTTTGG	2640
	GTAAGTAGTA	CGGACTATTT	AGCTGATGTT	AGAATTTCTG	ATGAAATTTT	AAAACAAACA	2700
	TCTGATGTTT	ATCGTAAAAT	CAGAAATACA	PTAAGATTTA	TGTTAGGTAA	CATTAACGAT	2760
50	TTCAATCCTG	ACACAGATAG	CATTCCTGAA	TCAGAGTTAT	TAGAAGTGGA	TCGTTACTTG	2820
	CTAAATCGTT	TACGTGAATT	TACTGCAAGT	ACGATTAACA	ACTATGAAAA	CTTTGACTAC	2880
55	TTAAATATTT	ATCAAGAAGT	TCAAACTTTT	ATCAATGTTG	AGTTAAGTAA	TTTCTATTTG	2940

	CAAACAGTGT TATATCAAAT TTTAGTTGAT ATGACGAAGT TGTTAGCACC AATCTTAGTG	3060
5	CATACAGCTG AAGAAGTTTG GTCTCATACA CCACATGTTA AAGAAGAAAG TGTTCACTTA	3120
	GCAGACATGC CTAAAGTTGT AGAAGTAGAT CAAGCTTTAT TGGATAAATG GCCTACATTT	3180
	ATGAATTTAC GTGATGATGT GAACCGTGCA TTAGAACTG CTCGTAATGA AAAAGTTATT	3240
10	GGTAAATCAT TAGAAGCTAA AGTTACGATT GCTAGTAACG ATAAATTTAA TGCATCTGAA	3300
	TTCTTAACTT CATTTGATGC ATTACATCAA TTATTTATCG TGTCAACAAGT TAAAGTTGTA	3360
	GATAAGTTAG ACGATCAGGC AACAGCTTAT GAACATGGTG ATATTGTCAT CGAACATGCA	3420
15	GATGGTGAAA AATGTGAAAG ATGTTGGAAC TATTCAGAGG ATCTTGGTGC TGTGATGAA	3480
	TTGACGCATC TATGTCCACG ATGCCAACAA GTTGTAATAA CACTTGTATA ATTGAAATTG	3540
	TATAAAGTAC TCATACAGAT GATATAAATT AAAGCTCTCT TCATAATCAT GTTGTAGTTT	3600
20	TTGTTGACAT GATGAAGAGA GTTTTTTTGT GAATAAAAAA ATGACCAAGT TACCGGTCAT	3660
	ATATGTAAAA AATGTGCGAT TTAATAAAAT AAAAATTATT CAGGAATGGT ACAAATTCTC	3720
25	TGAGGCATAT AAATGCGTTA TAGTTGCTAT TCTCAATTAT GTTCGCGATA ATTTTAAGTA	3780
	AAAGTAAGCA CAGATATTGA ATTTGATAGG AGTTAATTGA ATGTATCATA ACAGTAACGC	3840
	AAACTTTGTC AATGGTATCA CTTTAAATGT GAGAGATAAG AATGAATTAA AGCCATTTTA	3900
30	TGAGGACATA TTAGGATTAA ATATTATAAA TGAGACATTA ACATCGATAC AATATGAAGT	3960
	AGGTCAAAAT AATCATGTCA TTACACTTGT TGAATTACAA AATGGACGTG AACCTTTAAT	4020
	GTCCGAAGCG GGAAGTTTC ATATCGCAAT TAACTACCT CAAATTAGTG ATTTAGCTAA	4080
35	TTTACTAATT CATTTAAGCG AATATGATAT TCCAGTTAAC GGAGGTATAC AGCCTGCTTC	4140
	GTTATCATT TTTTTGAAG ACCCGGAAGG AAACGGTTTT AAATTTTATG TTGATAAAGA	4200
	CGAAGCGCAA TGGACGAGGC AAAATAATTT AGTAAAAATT GATATTAGAC CATTAAATGT	4260
40	ACCGAGATTA GTGAGTCATG CAACAAAATT GTTATGGTTA GGTATTCCAG ATGACGCTAT	4320
	TATAGGTGCA TTGCATATTA AGACAATTCA TTTATCAGAG GTAAAAGAGT ACTACCTCGA	4380
45	TTATTTTGGG TTAGAGCAAT CGGCATATAT GGATGATTAT TCAATATTTT TAGCATCGAA	4440
	TGGCTATTAT CAACATTTGG CCATGAATGA TTGGGTATCA GCAACGAAAC GTGTAGAAAA	4500
	TTTTGATACG TATGGATTAG CAATTGTTGA CTTTCATTAT CCTGAAACAA CACATTTAAA	4560
50	TTTACAAGGT CCGGATGGTA TCTATTATCG CTTTAATCAT ATCGAAGTTG AAGATTAGTA	4620
	TATACTTTGA ATGGACGAAC CATATAATGA ATCGTTTTTA ATGATCTTTT TATACAAGTT	4680
55	ATGAAGGAGG CTGGGACATT AAGTTCTTAG GCAATGTAAA AAGCTGATTT CTATTAATTA	4740

TTTTCCTTAT ATTAATTGCC ATTAATACAA AACCTAGCTC TCGTTTAACT TTATTTATTC 4860  
 CTGGAAGTGA CATTGCGnGTG AACTCAAAAT nGCCTACTTh CTAAATTAC CAATATCTAT 4920

5 (2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 626 base pairs  
 (B) TYPE: nucleic acid  
 10 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

TGGATTGCCA TTACATGGAC AAGATTTAAC TGAATCAATT ACACCATATG AAGGTGGTAT 60  
 CGCTTTTGCA AGTAAACCAT TAATTGATGC TGATTTTATT GGTAAATCTG TATTAAAAGA 120  
 20 TCAAAAAGAA AATGGTGCAC CAAGAAGAAC AGTGGGATTA GAATTACTTG AAAAAGGAAT 180  
 TGCAAGAAGT GGTATGAAG TTATGGATTT AGATGGAAAT ATTATTGGAG AAGTAACTTC 240  
 AGGAACACAG TCTCCATCAT CAGGAAAATC AATTGCACTT GCAATGATAA AAAGAGATGA 300  
 25 GTTTGAAATG GGTAGAGAGT TGCTTGTTCA AGTTGCTAAG CGTCAATTAA AAGCGAAAAT 360  
 TGTTAAGAAA AATCAAATTG ATAAATAATT AAAAAGGGGT GTGCATTGTG AGTCATCGTT 420  
 ATATACCTTT AACTGAAAAA GACAAGCAAG AAATGTTACA AACAATTGGT GCAAAATCTA 480  
 30 TAGGAGAATT ATTGCGGTGAT GTACCAAGTG ACATTTTATT AAATAGAGAT TTAAATATTG 540  
 CTGAAGGCGA ACGGAGAACA ACGTTACTTA GAAGATThAA TCGCATTGCA AGCAAGAGTA 600  
 35 TCACTAGAGG AACGCGTACA TCGTTT 626

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1126 base pairs  
 40 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

nGGAAGTGGT GTATATATTT GTAATGAGTG TATTGAATTA TGCTCAGAAA TCGTCGAAGA 60  
 50 AGAATTAGCT CAAAACACTT CTGAAGCGAT GACAGAATTA CCTACTCCTA AAGAAATTAT 120  
 GGATCATTTA AACGAATATG TTATTGGTCA AGAAAAAGCT AAAAAATCTT TAGCTGTAGC 180  
 TGTTTATAAC CACTATAAGC GTATTCAACA ATTAGGACCA AAAGAAGATG ATGTTGAATT 240

55

AACCTTAGCC AAGACGTTGA ATGTACCATT TGCAATTGCA GATGCGACAA GTTTAACTGA 360  
 AGCTGGTTAT GTAGGCGATG ATGTTGAAAA TATCTTGTTG AGATTAATTC AAGCAGCTGA 420  
 5 CTTTGACATT GATAAAGCCG AAAAAGGTAT TATTTATGTA GATGAAATTG ATAAAATTGC 480  
 ACGTAAATCT GAAAACACAT CTATAACAGC TGACGTTTCA GGTGAAGGTG TTCAACAAGC 540  
 10 ATTGCTTAAA ATCTTAGAAG GTACGACTGC AAGTGTCCG CCACAAGGTG GACGCAAACA 600  
 TCCAAACCAA GAAATGATTC AAATTGATAC AACAAATATC TTATTTATTC TTGGTGGTGC 660  
 CTTTGATGGT ATTGAAGAAG TGATTAAGCG CCGTCTGGT GAAAAAGTTA TTGGTTTCTC 720  
 15 AAGCAATGAA GCTGATAAAT ATGACGAACA AGCATTATTA GCACAAATTC GCCCAGAAGA 780  
 TTTGCAAGCC TATGGTTTGA TTCCTGAATT TATCGGACGT GTGCCAATTG TAGCTAATTT 840  
 AGAAACATTA GATGTAAGTGC CGTTGAAAAA CATCTTAACG CAACCTAAAA ATGCACTTGT 900  
 20 GAAACAATAT ACTAAAATGC TGGAATTAGA TGATGTGGAT TTAGAGTTCA CTGAAGAAGC 960  
 TTTATCAGCA ATTAGTGAAA AAGCAATTGA AAGAAAAACA GGTGCGCGTG GTTTACGTTT 1020  
 AATCATAGAA GAATCGTTAA TCGATATTAT GTTTGATGTG COTTCTAACG AAAATGTAAC 1080  
 25 GAaGGTAGTT ATTACAGCAC AAACmATTAA TGrAGaACTG AACCAG 1126

(2) INFORMATION FOR SEQ ID NO: 29:

30 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 4392 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

ATTGACTTCT TAGCAATnAA TaTGAGTGAA GAACGTACTG TTGAAGTACC AGTTCAATTA 60  
 40 GTTGGTGAAG CAGTAGGCGC TAAAGAAGGC GCGTAGTTG AACCAACCATT ATTCAACTTA 120  
 GAAGTAACTG CTAATCCAGA CAATATTCCA GAAGCAATCG AAGTAGACAT TACTGAATTA 180  
 45 AACATTAACG ACAGCTTAAC TGTTGCTGAT GTTAAAGTAA CTGGCGACTT CAAAATCGAA 240  
 AACGATTCAG CTGAATCAGT AGTAACAGTA GTTGCTCCAA CTGAAGAACC AACTGAAGAA 300  
 GAAATCGAAG CTATGGAAGG CGAACAACAA ACTGAAGAAC CAGAAGTTGT TGGCGAAAGC 360  
 50 AAAGAAGACG AAGAAAAAAC TGAAGAGTAA TTTAATCTG TTACATTAAA GTTTTATAC 420  
 TTTGTTTAAAC AAGCACTGTG CTTATTTTAA TATAACCATG GTGCTTTTGT TGTATTATA 480  
 AAGCTTAATT AAACCTTATT ACTTTGTAAT AAAGTTTAAT TAATTTTAGT GAGTAAAAGA 540

55

	CTTACTAAGC TAAAGAATAA TGATAATTGA TGGCAATGGC GGAAAATGGA TGTGTGCATT	660
	ATAATAATAA ATGAAACAAT TATGTTGGAG GTAAACACGC ATGAAATGTA TTGTAGGTCT	720
5	AGGTAATATA GGTAAACGTT TTGAACCTAC AAGACATAAT ATCGGCTTTG AAGTCGTTGA	780
	TTATATTTTA GAGAAAAATA ATTTTTCATT AGATAAACAA AAGTTTAAAG GTGCATATAC	840
10	AATTGAACGA ATGAACGGCG ATAAAGTGTT ATTTATCGAA CCAATGACAA TGATGAATTT	900
	GTCAGGTGAA GCaGTTGCAC CGATTATGGA TTATTACAAT GTTAATCCAG AAGATTTAAT	960
	TGTCTTATAT GATGATTTAG ATTTAGAACA AGGACAAGTT CGCTTAAGAC AAAAAGGAAG	1020
15	TGCGGGCGGT CACAATGGTA TGAAATCAAT TATTTAAATG CTTGGTACAG ACCAATTTAA	1080
	ACGTATTCGT ATTGGTGTGG GAAGACCAAC GAATGGTATG ACGGTACCTG ATTATGTTTT	1140
	ACAACGCTTT TCAAATGATG AAATGGTAAC GATGGAAAAA GTTATCGAAC ACGCAGCACG	1200
20	CGCAATTGAA AAGTTTGTG AAACATCACG ATTTGACCAT GTTATGAATG AATTTAATGG	1260
	TGAAGTGAAA TAATGACAAT ATTGACAACG CTTATAAAAG AAGATAATCA TTTTCAAGAC	1320
25	CTTAATCAGG TATTTGGACA AGCAAACACA CTAGTAACTG GTCTTTCCCC GTCAGCTAAA	1380
	GTGACGATGA TTGCTGAAAA ATATGCACAA AGTAATCAAC AGTTATTATT AATTACCAAT	1440
	AATTTATACC AAGCAGATAA ATTAGAAACA GATTTACTTC AATTTATAGA TGCTGAAGAA	1500
30	TTGTATAAGT ATCCTGTGCA AGATATTATG ACCGAAGAGT TTTCAACACA AAGCCCTCAA	1560
	CTGATGAGTG AACGTATTAG AACTTTAACT GCGTTAGCTC AAGGTAAGAA AGGGTTATTT	1620
	ATCGTTCCTT TAAATGGTTT GAAAAAGTGG TTAACCTCTG TTGAAATGTG GCAAAATCAC	1680
35	CAAATGACAT TGCCTGTTGG TGAGGATATC GATGTGGACC AATTTCTTAA CAAATTAGTT	1740
	AATATGGGGT ACAAACGGGA ATCCGTGGTA TCGCATATTG GTGAATTCTC ATTGCGAGGA	1800
	GGTATTATCG ATATCTTTCC GCTAATTGGG GAACCAATCA GAATTGAGCT ATTTGATACC	1860
40	GAAATTGATT CTATTCGGGA TTTTGATGTT GAAACGCAGC GTTCCAAAGA TAATGTTGAA	1920
	GAAGTCGATA TCACAACCTGC AAGTGATTAT ATCATTACTG AAGAAGTGAT CAGCCATCTT	1980
45	AAAGAAGAGT TAAAACTGC ATATGAAAAT ACAAGACCCA AAATAGATAA ATCAGTGCGC	2040
	AATGATTTGA AAGAAACGTA TGAAAGCTTT AAATTATTCG AAAGTACATA CTTTGATCAT	2100
	CAAATACTAC GTCGCTTAGT AGCGTTTATG TATGAAACAC CTTGACAAT TATTGAGTAT	2160
50	TTCCAAAAAG ATGCAATCAT TGCAATTGAT GAATTTAATC GTATTAAAGA AACTGAAGAA	2220
	AGTTTAACAG TAGAGTCTGA TTCGTTTATT AGCAATATTA TTGAAAGTGG TAATGGATTT	2280
55	ATAGGACAAA GTTTTATAAA ATATGATGAT TTTGAAACAT TGATTGAAGG CTATCCTGTC	2340

	TCATGTAAAC CTGTCCAACA ATTTTATGGG CAATATGACA TTATGCGTTC TGAATTTCAA	2460
	CGATATGTTA ATCAAACTA TCATATCGTG GTTTTGGTCG AAACCGAAAC TAAAGTTGAA	2520
5	CGTATGCAAG CGATGTTAAG TGAATGCAT ATTCCATCAA TAACAAAATT GCATCGCTCA	2580
	ATGTCATCGG GGCAAGCAGT GATTATTGAA GGCAGTTTAT CTGAAGGATT TGAACCTACCT	2640
10	GATATGGGAT TAGTTGTCAT TACTGAGCGT GAGcTTTTTA AATCAAAACA GAAAAAGCAA	2700
	CGAAAACGTA CGAAAGCTAT CTCAAATGCT GAAAAAATTA AGTCTTACCA AGATTTAAAT	2760
	GTGGGAGATT ATATTGTTCA TGTGCATCAT GGTGTTGGTA GATATTTAGG TGTGAGACG	2820
15	CTCGAAGTGG GGCAAACGCA TCGTGATTAT ATTAAATTGC AATATAAAGG TACGGATCAA	2880
	CTATTTGTTC CAGTAGATCA AATGGATCAA GTTCAAAAAT ATGTAGCTTC GGAAGATAAG	2940
	ACGCCAAAAT TAAATAAACT CGGTGGCAGT GAATGGAAAA AAACAAAAGC TAAAGTTCAA	3000
20	CAAAGTGTTG AAGATATTGC TGAAGAGTTG ATTGATTTAT ATAAAGAAAG AGAAATGGCA	3060
	GAAGGTTATC AATATGGGGA AGACACAGCT GAGCAAACAA CATTGGAATT AGATTTTCCA	3120
25	TATGAACTTA CGCCTGACCA AGCTAAATCT ATCGATGAAA TTAAAGATGA CATGCAAAAA	3180
	TCGCGTCCAA TGGATCGCTT GCTATGTGGT GATGTTGGTT ATGGTAAAC TGAAGTTGCA	3240
	GTGAGAGCAG CATTCAAAGC TGTAATGGAA GGAAAGCAGG TTGCATTTTT AGTTCCTACA	3300
30	ACTATTTTAG CTCAGCAACA TTATGAGACG TTAATTGAGC GTATGCAAGA TTTTCCTGTT	3360
	GAAATTC AAT TAATGAGTCG TTTTAGAACG CCTAAAGAGA TAAAACAAAC TAAGGAAGGA	3420
	CTTAAACTG GATTTGTTGA CATAGTTGTT GGTACACACA AATTACTTAG TAAAGATATA	3480
35	CAGTATAAAG ATTTAGGGCT GTTGATTGTA GATGAAGAAC AACGATTTGG TGTACGCCAT	3540
	AAAGAGCGTA TTAAACATT AAAACATAAT GTAGATGTAC TAACATTGAC TGCAACCCCA	3600
	ATAGCTAGAA CATTGCATAT GAGTATGCTA GGTGTGCGGG ATTTGTCAGT GATTGAAACG	3660
40	CCGCCAGAAA ATCGTTTCCC AGTTCAAACA TATGTATTAG AACAGAACAT GAGTTTTATC	3720
	AAAGAAGCTT TAGAAAGAGA ACTATCCCGT GATGGCCAAG TGTTTTATCT TTATAATAAA	3780
45	GTGCAATCCA TTTATGAAAA ACGAGAACAA CTCCAGATGT TAATGCCAGA TGCTAACATT	3840
	GCAGTTGCTC ATGGACAAAT GACAGAGCGC GATTTAGAAG AAACGATGTT AAGTTTTATC	3900
	AATAATgAAT ATGATATTTT AGTAACGACG ACGATTATTG AAACAGGTGT CGATGTCCCA	3960
50	AATGCAAATA CTTTGATCAT TGAAGATGCA GATCGCTTTG GATTGAGTCA GTTGATCAA	4020
	TTAAGAGGTC GTGTTGGTCG TTCAAGTCGT ATTGGTTATG CATACTTCTT ACATCCAGCA	4080
55	AATAAGGTAC TAACTGAGAC TGCAGAAGAT CGATTACAAG CGATTAAAGA ATTTACGGAG	4140



TTAGGTAAAC AACAGCACGG CTTTATTGAT ACAGTTGGAT TTGATTGTGA CAGTCAAATG 4260  
 TTAGAAGAAG CTGTAAATGA AAAACGTGGT ATTAAGGAAC CAGAATCTGA GGTGCCAGAA 4320  
 5 GTCGAAGTTG ATTTAAACTT GGATGCATAT TTGCCAACAG AATATATTGC AAATGAACAA 4380  
 GCTAAAATTG AA 4392

(2) INFORMATION FOR SEQ ID NO: 30:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 729 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 15 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

20 TTTCTTTTGA ATCTATATCG AGGTGGTTGG TAGGTTTCATC TAAAATAAGT ACATTGTCAC 60  
 GTTGCAACAT AAGTAGTGCT AGTTGTAAAC GTGCTTTTTC ACCACCAGAT AAATCATTTAA 120  
 25 TTATCTTTTT AACATCGTCT TGTACAAATA AGAAACGTCC AAGAACTGCT CGAATATCTT 180  
 TTTCAATTCAT TAACGGATAT TGATCCCACA CATAATCTAA AATCGTTTTA CTAGATTTAA 240  
 ATTCTGCTTG CTTTTGATCA TAATAACCAA TTTGTAAATT TGCGCCGAAA GTAATATCGC 300  
 30 CATTAAAGCGC TTTTGTGTTGA TTAGCAATAG TTTTAATTAA GGTCGATTTT CCAATACCAT 360  
 TTGCCCCAAT GATTGCTATA TGATCGCCTT TAGAGACCTC TATACTCATA GGTTTGGTAA 420  
 TTGCAGTTTG ATAACCGATT TCTAAATTTT TTACATGCAT GACGTCATTA CCTGTATTC 480  
 35 GGTCAAAGCC AAATTGAATA TTTGCACTTT TGGCATCTAA CATTGGTTTA TCAATGCGTT 540  
 CCATTTTTTC TAAAATCTTA CGTCTACTTT TTGCCATTCC ACTTGTGTA GCACGGGTAA 600  
 TATTTTCTC AACAAAAGTT TCTAATCGTT TTATTTCTGC TTGTTGACTT TCATATTCTT 660  
 40 GCATTCGTTT TTGATAATAT AAATCCCGTT GCTGTATAAA TTCCTCGTAA TTACCAACAT 720  
 AGCGTTTGA 729

(2) INFORMATION FOR SEQ ID NO: 31:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 13856 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 50 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

	TGATGTTTCG ATACATTTGT TGCACCTTGT GGATATACTT TAAAGGTTGT GTCGTATGTT	120
	TCCTTACTAT CTTTAGCTTC AGATTCCTGT GATTCAACCG TTTTATATTT TTCAAGTGCA	180
5	TGTCCTTCAA TATCAACTCG TGAATAATG CGATTCAACC ATGCTGGTAA ATACCACGAA	240
	CCTTTtCCAA ACAATTTcGt TAATGCAGGA ATTAACATCA TtCTGACTAC GAAGGCATCA	300
10	AAGAGTACAC CAAACGCTAA TGCCATACCC ATTGATTAA TCATGACATC TTCTTGGAAT	360
	ACAAACGCAA AGAAGACACT AAACATAATT AATGCAGCTG CTACAATAAC AGGACCGCTT	420
	TCTTTCAATC CTACTTTGAT AGAATAATCA TTATCCCCTG TTTTACTATm yyCTTCATGr	480
15	ATTGCGGACA TAAGGAAGAC TTCATAATCC ATCGCTAATC CAAATAAGAT ACCTATAGTA	540
	ATAACCGGTA AAAATGCTAG CATTGGTCCT GTCGTTTCAA TACCAAACAG ACCTTTCATA	600
	AAACCATCTT GCATTACTAA TGTGTAAAT CCTAATGTTG CCATTAAATGA CAAGACGAAT	660
20	CCTAAACTG CTTTTAATGG TATTAGAATT GAACGGAAGA CAATCATTAA TAAGAAAAAT	720
	GCTAATACAA CAATGACTGA GGCAAATAAA GGTATCGCCT CATTTAACTT TTTAGACATA	780
25	TCAATATTAA TGACACTTTG TCCCGAAATC TCCGTTTGA ACCCATATTT ATCTTGTGCA	840
	TCTTTATGAT AATCTCGTAA ATCATGCACT AAATCATTTG TACTCTCTGC ATTAGGCCCT	900
	TGCTTAGGTA TCACGACCAT CAAAGCGTAA TCATTATCTT TACTCATTTG TGGTGGCGTA	960
30	ACGATATCTA CATTTTTCTT ATCTTTAATA TCTTTATATA CAGACTGTAA ATCTTGTGT	1020
	AATCCTGTG GATCATCCTT TTTATCTTTC ACATTATCA ACATCGGTAT TTGGCCATTA	1080
	AATCCTTCAC CAAATTTATC CGAGATAATA TCGTAAGCTT TTTTCTGTGT AGAATCTGCT	1140
35	GGTTTAACAC CGTCATCTGG AATACCAAGT CGCATATGAC TAACTGGTAT TGCAGCTGCT	1200
	ACTAATATGA TTAAACCTAG TAATACTGCC GCAAGTGCAT TTCCTGTAAT AAATTTAGAC	1260
	CATGGCGTAT CAATATCTTT TTTGAATTTA GACTGTAATT TATTCACTTT AATGCGTTtA	1320
40	TGGAAAATGC TTATTAATGC AGGTAATAAA GTTAAAGCGC TAAGTACTGC AAAACAACA	1380
	CTAATTGCCG AAGCAAATCC CATTACCGCT AAGAAGTCAA TGCCTACTAA TGATAAACCA	1440
45	CATACTGCAA TTACAACTGT TACACCAGCA AAAACAACTG CACTACCTGC TGTTCCTATT	1500
	GCAAGACCAA TGCCTTTAAT GTAATCTGTT TCAGTTTTCA TAACTTGTG ATATCTGAAT	1560
	AAAAATAATA ATGCATAATC GATACCAACT GCTAGTCCAA TCATTACGGC TAATGTCAGT	1620
50	GTGACATTTG GTATATCGAA TGCATAAGTT AACAACTGA TAATACCTAC ACCAGAGGCT	1680
	AGACCAATCA ATGCACTTAT AATTGGTAAT CCTGCAGCAA TGAAGTGAACC GAATGTGATT	1740
55	AACAGTACAA CAAATGCAAC AATAATACCA ACTAGTTCAG AATTACCGCC TACTTCTGTA	1800

	AAATGACTTT TAACATTATC TCTAGAGCCA TCTTTTAAAG ATGTTTGACT AACGTCATAT	1920
	GTGATATCTG CAAATGCAGT TGTTTTATCT TTAATAATTT GCTTATTTTC ATAAGGATCT	1980
5	GATATTTTAT CAATGTGCTT GTCATCTTTT TTAATATCAT CTAACGTTTT CTTAATATCT	2040
	TTAGTAATGT TCGGTTCAC AATACCATCA TCTTTAGTCG TCTTAAAGAC AACACGTATT	2100
10	TGTGCCTTTT CACTATCTTG ATTAATAATGT TTTTCAATCT TTTTATTCGT ATCTAACGAC	2160
	TCTAATCCTG TCATTTTAAT ATCATTGTCA AATTTTCGGTG CATTTGTAGC AAGTGGTATC	2220
	AATATTGCAG CTACAATCAC TATCCATGCA ATGACCGCGG ACCATTTATG TTTTGCGATG	2280
15	AATGTCCCCA TCTTATATAA AAATTTTGCC AAAGTATATT GCCTCCTTTT AAAATCAACG	2340
	TTATAGTTTA AATATACAGT GTAGATTATT GTTCGATTAT AGTATCTATC CCCGACCTCT	2400
	TAAAGAATCA ATTGGAAAAT TTTGTATATT AAACACACA CAAAGGAGAA ATGTAGATGA	2460
20	AAGAGACTGA TTTACGAGTT ATAAAGACAA AAAAAGCATT GTCGAGTAGC TTGCTACAAT	2520
	TGTTAGAACA GCAATTATTC CAAACGATTA CTGTCAATCA AATTTGCGAC AACGCACTCG	2580
25	TACACCGTAC AACATTTTAT AAACATTTTT ATGATAAATA TGATCTTCTA GAGTACTTGT	2640
	TCAATCAATT GACTAAAGAC TACTTTGCTA GAGATATCAG TGACCGTCTT AATCATCCAT	2700
	TCCAAACGAT GAGTGATACG ATTAATAATA AAGAGGATTT GAGAGAAATC GCAGAATTCC	2760
30	AAGAAGAAGA CGCTGAATTT AATAAAGTAT TAAAAAATGT CTGCATTAAA ATTATGCATA	2820
	ACGATATCAA AAATAATAGA GACCGTATCG ATATTGACAG CGACATCCCA GATAATCTCA	2880
	TATTTTATAT TTATGACTCG TTGATTGAAG GTTTTATACA TTGGATAAAA GATGAAAAAA	2940
35	TTGATTGGCC TGGCGAAGAT ATTGATAACA TTTTCCATAG ATTAATCAAT ATTAAGATTA	3000
	AATAGTAGAT GAGAACTCA TGAGCGTTAC CAACATTCAT AATAAAAACG ATAGTGKACA	3060
	CGTTAATGAA TTCGTGTACT ACTATCGTTT TTTATTTTTA TCGTGCTTAT CGCTATTAAA	3120
40	ACAACTGATA CACAACACAT AAACATGAA GAAAAAATA AATCCGCTAT CTAAATGACT	3180
	TTGACTCAGT TGTTTAAATG ACCAAATTGC TAATACAATT CCCATTATTA TTGAAATAAC	3240
45	GTATCTCACA TTCTTATACC TATAATCCTT TTCTAAAAAT ATGGTTGCTA TTAATAATT	3300
	TTTAAAGTTA TAAATAAAAA GAGCCAACCG CAATGGATGG CCCTTGTTCA TTATGAAGCA	3360
	TTAGAACATT TCTGAAACAA CCTTTTGTTT TAAGAAGTGT AATAAGTAGT CTGGACTACC	3420
50	TGTTTTAGCG TCCGTACCTG ACATTTTGAA ACCACCAAAT GGATGGTATC CAACAACTGC	3480
	TGAAGTACAG CCTCTGTTAA GGTATAAATT GCCTACATCA AATTCGTTTA CCGCTTTAAT	3540
55	CCAATGCTCG CGATTATTTG TAATCACTGC ACCAGTTAAA CCGTAATCTG TATCATTTGC	3600

	TTCTTCTTGC ATGATTCTAT CTTTAGATTT AAGTCCTGAA ATGATTGTTG GTTCTACAAA	3720
	GTAACCTTTT GAATCATCAG TGCCGCCACC TTGTTCTAAT TTACCTTCTT CTTTACCAAT	3780
5	CTCAATATAA TTTTAAATCT TATCAAATTG TTTTTTATTA ATAAC TGGGC CCATATACGT	3840
	ATTGTCTACA GTATTGCCCA ACGTTAATTC TTTTGTAAAT TTGATTGATT TCTCTAATAC	3900
10	TTCGTACATAA ACGTCTTTAT GCACAATTGC ACGTGAACAT GCTGAACATT TTTGACCAGA	3960
	AAAACCAAAT GCTGACGTTA CAATAGCTTC TGCTGCCATA TCTGTATCAA TATTTTCATC	4020
	AACTACAATG GCATCTTTAC CACCCATTTT AGCGATAACA CGTTTCAAGA AGTTTTGACC	4080
15	TTCTTGAACA ACGGCACTAC GTTCATAAAT TCTAGTACCT GTCGCACGTG ATCCTGTAAA	4140
	TGTAACGAAA TCGGTATCTT TATGATCAAC TAAGTAATCA CCAATTTCTT TCGGATCACC	4200
	AGGAACAAAG TTAAC TACGC CTTTTGGTAA TCCTGCTTCT TCTAAAATTT CCATTAATTT	4260
20	ATAAGCGATA TAAGGTGTAT CCTCAGCAGG TTTCAATAAC ACTGTATTAC CTGCCACAAC	4320
	TGGTGCTAAA GTTGTACCAG CCATAATCGC AAACGGGAAG TTCCACGGCG GAATTGTAAC	4380
	ACCTGTACCA ATTGATTTAT AGAAATATTT ATTGTGTTCA CCTTCACGAT CAAGTACTGG	4440
25	CTTACCTTGA GCCAAGTCCA TCAITGAACG TGCATAGTAT TCAATAAAAT CAATACCTTC	4500
	AGCTGCATCA CCAACTGCTT CATCCCATGG CTTACCTGCT TCATAAACCA TAATTGCTGC	4560
30	AATTTCCGCT TTTTCGACGAC GAATAATTGC CGAAACACGT AACATAAGCT CTGCACGATC	4620
	ATTTGCTGAC CATGTTTTCC AAGATTTATA AGCTTCGTTT GCTGCTTTAA ACGCATCTTC	4680
	AACATCTTGT TTTGTTGCCT TTGATGCATT TGCAATCACT TGTGATGTGT CTGCAGGATT	4740
35	GATTGATTTA ATTTTGT CAT CTTTGAAAAT CTTCTCTCCA TTAATCACTA ATGGTATGTC	4800
	TTGACCTAAT TCTTTTCCA CGTCTTCAA TGCTTTCTTA AACATATCCA CATTTTCTTG	4860
	GACTGAAAAA TCGTAACCAG GTTCATTTTT AAATTCTACT ACCATGTACA CTTACCCCTT	4920
40	ATAAATTTTG AAAGTGGTTT AACCCTTTGA TTAAATGATA TAACATCATT TAAACTCATT	4980
	TTACTATGAT TAAGGTIAGT TTTGCAATCG CTTTCATTTT TATGTTTTAT CACTTATTCT	5040
45	CAAGTATTTT GAAATTGATT GGTACTTTT TAAAATTTAT ATGGGTCGCA ACTGCTACTT	5100
	TATCGTTTCG TCATTTAATG TTTCCGATGG TAGGTCATTA TCAATTTTAC GAACGACTTT	5160
	ACAAGGGTTT CCAACCGCTA AGCTGTGTGG CGGAATATCT TTAGTGACAA CACTACCAGC	5220
50	ACCAATCACA CTGCCTTCTC CAATCGTCAC CCCTGGTAAC ACGGCTACAT GACCGCCAAA	5280
	CCAAGTATTA CTGCCAATAT GAATGGGTCC GGCTTTTCA AAACCTTCAT TTCTATGATG	5340
55	GAAATTAAGT GGATGTGTCG CTGTGTAGAA TCCACAATTA GGTCTATATA AACATTATC	5400

	TCCTAGTTTA	ACGTTCCAAC	CATAATCTGT	ATCAAAAGGA	ATCGAAATAC	TTACATTGTC	5520
	TGTTGTTGTT	TGAAATAATT	GATCAATTAA	TTCTTTCTT	TTATTTGTAG	CACTCGGTCT	5580
5	TGTATGATTT	AATTCAAAGC	AAATATCTTT	CGCTCGTGCA	CGTTCATTGA	TTAAGTATTG	5640
	ATCAAAGTTT	GCATCGTACC	ATTTTTCTGC	TAACATTTTT	TCTTTTTCAG	TCATTACACC	5700
10	TTTCAACTCC	TAATAACTTA	TTTACTTGTT	TAAAAGTTAA	TCAAATAAAC	CTTCGCCTAT	5760
	GCAACTAATA	CGCTATAACA	TTATGAAATC	ATGACCTTAT	CACCCTTATC	TATACAATTC	5820
	TCGCATCAAA	TACTGCTAAA	GTAGTAGATA	AATTCAATAC	TACAGACGCA	TTCATTTTTT	5880
15	AATCTATTAA	CGTACAATGT	GAGTAAGAGA	AATATAAAGG	AGTATGATAG	CGATGAGAAT	5940
	ATTAATTACA	GGCACAGTTG	CTATCTTAAT	CATTCTAGGT	TTGGTCAAAA	CGATACAAGA	6000
	TTACGAAATG	ACAAACGACA	CGAGTCGTcA	GTTGTCAGAC	AACAAAGATG	ATGATAAAGT	6060
20	CATCCATCTT	AATAATTTTA	AAAATTTACA	TGCGAAAGAA	TTTAACCCAT	CTGATTTCTT	6120
	TTAAGTCACC	TAAGAATTGC	AAATCCAGAA	GTCATTTAAG	TTTTACCTTT	CATTCATACA	6180
25	TCCTTTAATA	TTAATTACGA	CTTCTTTTAT	ATAGATGCTA	AGTAGAGAGA	TTGTTGTGCA	6240
	ATGTTTGAC	GGCAATCTCT	CTTTTCTTTT	TTAAAATTGG	TAAAAGTAAA	ACGCAACGAT	6300
	TGACTTATAT	ACCTATAGGG	GGTACATTAG	ACGTGTAACA	ATGAATCACA	GGGAGGCAAT	6360
30	AATGTGGCTA	ATACGAAAAA	AACAACATTA	GATATCACTG	GTATGACTTG	TGCCGCATGT	6420
	TCAAATCGTA	TCGAAAAGAA	ACTGAATAAA	CTTGATGACG	TTAATGCCCA	AGTGAATTTA	6480
	ACTACAGAGA	AAGCAACTGT	TGAGTATAAC	CCTGATCAAC	ATGATGTCCA	AGAATTTATT	6540
35	AATACGATTC	AACATTTAGG	TTACGGTGTC	GCTGTAGAAA	CTGTGCAATT	AGACATTACA	6600
	GGTATGACTT	GTGCTGCATG	CTCAAGCCGT	ATTGAAAAAG	TGTTAAATAA	AATGGACGGC	6660
	GTTCAAAATG	CAACGGTCAA	TTTAACAACA	GAGCAAGCTA	AAGTTGACTA	TTATCCTGAA	6720
40	GAAACAGATG	CTGATAAACT	TGTCACTCGC	ATTCAAAAAT	TAGGTTATGA	CGCGTCTATT	6780
	AAAGATAACA	ATAAAGATCA	AACGTCACGC	AAAGCTGAAG	CGCTACAACA	TAAATTGATT	6840
45	AAGCTTATCA	TATCAGCAGT	ATTATCTTTA	CCACTATTAA	TGTTAATGTT	TGTACATCTT	6900
	TTCAATATGC	ATATACCAGC	ACTATTTACG	AATCCATGGT	TCCAATTTAT	TTTAGCTACA	6960
	CCTGTACAAT	TTATTATTGG	ATGGCAATTT	TATGTAGGTG	CTTATAAAAA	CTTAAGAAAT	7020
50	GGTGGCGCCA	ATATGGATGT	ACTTGTTGCT	GTTGGTACAA	GTGCAGCATA	TTTTTACAGT	7080
	ATTTATGAAA	TGGTTCGTTG	GCTAAATGGC	TCAACAACGC	AACCGCATTT	ATACTTTGAA	7140
55	ACAAGCGCCG	TACTAATTAC	CTTAATCTTA	TTCGGTAAGT	ATTTAGAAGC	TAGAGCGAAG	7200

	TTAAAAGATG GTAATGAAGT GATGATTCCT CTAAATGAAG TACATGTTGG AGATACACTT	7320
	ATCGTTAAAC CAGGTGAAAA GATACCTGTT GATGGCAAAA TTATTAAAGG TATGACTGCC	7380
5	ATCGACGAAT CTATGTTAAC AGGTGAATCT ATCCCTGTTG AGAAGAATGT TGATGATACT	7440
	GTAATTGGTT CAACGATGAA CAAAAACGGT ACTATTACTA TGACAGCAAC AAAAGTTGGC	7500
10	GGGGACACTG CGTTGGCAAA TATTATTAAA GTTGTCSAAG AAGCTCAAAG TTCTAAAGCG	7560
	CCGATTCAAC GATTGGCAGA TATTATTCTT GGTATTTCG TTCCTATCGT TGTGTTATC	7620
	GCACTATTAA CATTTATCGT GTGGATTACT TTAGTTACAC CAGGTACATT TGAACCTGCA	7680
15	CTTGTTGCGA GTATTTCGGT TCTCGTCATT GCTTGTCCAT GCGCATTGGG ACTTGCTACA	7740
	CCAACTTCTA TTATGGTAGG TACTGGTCGC GCTGCTGAAA ATGGTATTTT ATTTAAAGGT	7800
	GGCGAGTTTG TTGAACGCAC ACATCAAATT GATACCATCG TTTTAGATAA GACGGGTACC	7860
20	ATTACAAATG GTCGTCCAGT CGTGACAGAT TATCATGGTG ACAATCAAAC GCTACAACATA	7920
	CTTGCTACTG CTGAAAAAGA TTCTGAACAC CCATTGGCAG AAGCCATTGT CAATTATGCA	7980
25	AAAGAAAAGC AATTAATATT AACTGAGACA ACAACATTTA AAGCAGTACC TGGCCATGGT	8040
	ATTGAAGCAA CGATTGATCA TCACCATATA TTGGTTGGTA ACCGTAAATT AATGGCTGAC	8100
	AATGATATTA GCTTGCCTAA GCATATTTCT GATGATTTAA CACATTATGA ACGAGATGGT	8160
30	AAAAGTCTA TGCTCATTGC TGTAAATTAT TCATTAACTG GTATCATCGC AGTGGCAGAT	8220
	ACTGTCAAAG ATCATGCCAA AGATGCTATA AAACAATTGC ATGATATGGG CATTGAAGTT	8280
	GCCATGTTAA CTGGCGATAA TAAAAACACT GCTCAAGCCA TTGCAAAACA AGTAGGCATA	8340
35	GATACTGTTA TTGCAGATAT TTTACCAGAA GAAAAAGCTG CACAAATTGC GAAACTACAG	8400
	CAACAAGGTA AGAAGGTTGC GATGGTTGGT GACGGTGTA ATGATGCACC TGCATTAGTT	8460
	AAAGCTGATA TCGGTATCGC CATTGGTACA GGTACAGAAG TTGCCATTGA AGCAGCTGAT	8520
40	ATTACTATTC TTGGTGGCGA CTTGATGCTT ATTCCTAAAG CCATTATATG AAGTAAAGCA	8580
	ACCATTCGTA ATATTCGTCA AAATCTATTT TGGGCATTCT GCTATAATAT TGCCGGTATC	8640
45	CCTATAGCTG CATTGGGCTT ACTTGCGCCA TGGGTTGCTG GTGCTGCAAT GGCACTAAGT	8700
	TCAGTAAGTG TTGTCACAAA CGCACTTAGA TTGAAAAAGA TGCGATTAGA ACCACGCCGT	8760
	AAAGATGCCT AGATTCCTTA ATAATGAAGG ATTCGTTGGT GATTCTGAGA TAGGCTAGTG	8820
50	ATTGGCTCTA TAATGTCGCG GTTTAyaGt GGATCTTCGC TCCAAC TGCA TATATAGTnA	8880
	CACTTTTCGC TTGGCGAATT AGTGATCTT ACCTAATAGc TCCGCCTATT AGGTTCCATC	8940
55	ATTATTATAA ATAATAAGTA CACTACGGtT TACAGTTGGA TCTTCGCTCC AACTGCATAA	9000

	GAAATTTTAA ATGTTGAAGG TATGAGCTGT GGTCACTGCA AAAGTGCTGT TGAATCTGCA	9120
	TTAAATAATA TTGACGGTGT CACTTCAGCT GACGTTAACC TTGAAAATGG TCAAGTAAGT	9180
5	GTTCAATATG ATGACAGTAA AGTTGCTGTA TCTCAAATGA AAGACGCAAT TGAAGATCAA	9240
	GGTTACGATG TCGTTTAAAT AGGCAATATT CAACGTCATC AACACCAAAT TAAAAAATCG	9300
10	AACTGATGAG AATCCCAACA ATCCAAATTA TCTCATCAGT TCGATTTTTA ATTTACTCGT	9360
	AACCTAGTAT CTCCAGTCTG CAATACATCT AATGTTGCAT CTAATGCATC GACAATTAGA	9420
	TTTTTAACTG CAGCTTCAGT ATAAAACGCA ATATGTGGTG TTAATATGAC ATCTTCCCTG	9480
15	TCAATCAACG ATTCTAACAA TGGATCGTTC AGTGTTTTGC CCCTTTGATC ACTTGGGAAA	9540
	AGTTTGC GTT CAAATTCATA CGTATCAAGT GCTGCACCTT TAATCACACC ATTGTCTAAT	9600
	GCGTCTAATA ACGCCTTAGT ATCTACTAAA GAACCTCTCG CACAATTGAC AAATACTGCG	9660
20	CCCTTTTAA AATGTTTAAA TAATTCAGCA TTAAATAGAT AATGATTATA TTTCGTTGCA	9720
	GGTACATGTA ATGTCACGAT ATCAGCACCT TCAACCGCTT CCTCAATCGT ATCTTTGTAA	9780
	TCGACATACG TTGCAATTTT AGCATTAGGA AACGGTCGTA TGCGACCACA TCACTTTGAT	9840
25	AACCATTGGC AAATATATCG GCTACTACAC GGCCAATTG ACCTGTACCA ATAACAGCTA	9900
	CTTTTAAATC TTTAATGGAT TTCGATAAAA TAGTAGGTTT CCATCTAAAA TCATGCTCCC	9960
30	GCACTTTCGT TTGAATTTGA TTAAAATGAC GAACCACATT AATAGCCTGG TTCACAGCAA	10020
	ACTCCGCAAT TGAATTCGGA GAGTATGACG GCACATTTGA CACAATAAAG TTATACTTGT	10080
	TTGCTAACTC CAAATCATAT GTATCAAATC CAGCACTACG TTGTGCGATT TGTTTAATAC	10140
35	CTAGTTCATT TAATCGTTTA TAAACATGCT CTGATAATGG TATTTGTTGT GATAGCGATA	10200
	AGCCATCATA ACCAGCGACA CCTTCAACAT TGTCATCAGT TAATGCTTCT TTAGTAATAT	10260
	CTACCTCAAC ATGATGTTTC TCTGCCCACG CCTTGATATA AGGCATATCT TCATCACGTA	10320
40	CACTCATGAT TTAAATTTTT GTCATTTTAA CATCACCCTT AACTTTATTA TTCATATAAA	10380
	TATGCTAGTT CTGTTAATCT TATTGCAGCT TCGTCTAATT TCTGGTCATC TAACGCCAAT	10440
45	GAAATTCTCA CATAACGATT ACCATTCTCT CCAAATGGTT TCCCTGGAGC AACAAGTATT	10500
	GACTTCTCTT GCACTAAAAA TTGCTCAAAT TGCTCGCTGT CATAACCAGG CGGTGTTTCC	10560
	AACCATACAT ATATGCCACC TTTAGCATGA ACAAATGGCA AATCAGCTTT TGCAAGCATG	10620
50	GCTTCGAATC GGTCACGACG TGTTTTAAAT ACATTGCTTT GTTCTTCTAA AAAATCATCA	10680
	TAATGATTCA AAGCATATAT TGCGGCATCT TGTAATGCAC CAAACATCCC AGCATTTGTG	10740
55	TGCGTTTGGT ACTTTTTCAA AGCTTGAATC ATATCTTTAT TACCAACTGC AAAACCGACT	10800

	CCATTTTCCG	AAGCAAGTAT	ACTAGGATTT	TTAGCGTCGA	AACCGAAAGC	ACCATAAGCA	10920
	AAATCATGCA	CGATTTTAGT	GTCTGTACCT	TTAAATTTAG	CTATCGCTTC	ATCAAAAAC	10980
5	TCTTTCGTAG	CTGTGCATCC	AGTTGGATTA	TTTGGATACG	TTAAATAAAT	GAGTTTTGTT	11040
	TTATCTATTA	TTTGTGAATC	AACTTTGGAC	CAATCTGGCA	AATAATGTGG	CGGTTCTAAA	11100
10	TTAAGCGGGA	CTGGCTTGCC	ATCAGCTAAA	AGTACACCTG	CTAAATAATC	CGTGTAGCCT	11160
	GGATCAGGTA	GTAATACATA	GTCTCCTGGA	TTGATAACAC	ATGTTGGTAC	TGCCACTAAT	11220
	CCATTTTTTG	TACCATATAA	AATGCATACT	TCATCTTCTT	TATCTAACGT	CACATTATAT	11280
15	TGTCTTTGAT	AAAAATCTAC	AATAGCTTGC	TTGAACGCTT	CTTTACCATG	AAAAGCACCA	11340
	TATTTTTGAT	TTTCAGGAAT	AGTTAGTGCT	TTTTGAAAAT	GATCAATAAT	ACCTTGTGGC	11400
	GTGGGCCCCAT	CAGGGATTCC	AACTGCCATA	TTAATTAATG	GCAATGGTCC	ATGTTGATT	11460
20	TTACGTCCCA	TCGTTTTCCC	GAAATAACTA	TCAGGGATAT	TTGCTAATTT	GTTAGAGATC	11520
	ATCAAATTC	TCCTCTATCA	TTAAACATAG	CCTGGGCGAC	TATCATAATC	CTAACAACCT	11580
25	GTATCACTCT	CATTTAGATG	GTTACAATGA	CATCGCCATT	CACCGTTATG	TTCAACAGAA	11640
	CTTATGACAC	ACGTTGTATT	GAATGAATTT	ATTTTCATTT	TAGGTAGGTA	TAATATTATT	11700
	GTCAATATTA	GGAATTTTCA	GATTAATATG	CACTCAATCG	TTATGATTTA	ACTGTCATGC	11760
30	ATATCCGCAT	GCGCAACCAG	TTAGATATGC	TTATATAAAG	TATAACGCCC	ATCAAGGTAC	11820
	GTATTCAAAC	GTGAACCTTA	ACAGGCGTCA	TTCAATTGTTA	AATAAAACCT	CTTAAGCACA	11880
	TACTTATTTT	ACTATGCCTT	TTACGTTCCC	CTTATACTTT	TCTCACATCT	TTCTCTTAGA	11940
35	CTACTCCCTT	ATACGCCCCG	CTCAATATCT	TTAATCATTT	CATCTACAGT	TATTTTCGCA	12000
	CTCGTTAAGA	CAATAGGAAC	GCCTGCACCT	GGATGCGTAC	TTGCACCTGC	AAAATATAAA	12060
	TCTTTATAAT	CTCGCGATAC	ATTTTGTGGA	CGATAATAAT	TACTTTGCGC	TAAAGTTGGC	12120
40	ATTAAACCGA	ATGCCGAACC	AAATTTTCGCA	TGATACGTTT	GCTCAAAATC	ATTTGGCGTA	12180
	AAGATTGTTT	CTGAAACAAT	ATGCGATTTT	ATATCTTCAA	ATACTTCAAT	CGTTGCTAAT	12240
45	TTACGATAAA	TAATTTCCCT	TATTTGTTGC	GTCAAAGCTT	CATCTGACCA	ATCGATTCCG	12300
	CTACCTGTTT	TAAGTTCCGG	CGTCGGCATT	AGCACATAAA	TACCAGTTTT	GCCTTCTGGC	12360
	GCAAGTGATT	TATCAGCGAC	CGCTGGTACA	TACACATAAA	TAGAAGGATC	ATATGATAAA	12420
50	CGTCCCTCAA	ATATTTCTTC	AATATTGCCT	CTAAAGTCAT	CTGAAAAAAT	AACATTATGA	12480
	AGTCTCACTT	GATCTGTCAC	ATCAATATCT	ATACCGATAT	ACATTAAAAA	TGCTGAACAA	12540
55	GAGTAATCTA	AGTCTGCAAT	TTTATGTGGT	GGATACTTTT	TAATAGGTGC	AAAATCTGGC	12600



5 ATGTCACCAT TCACTTTTAT CGCATCGGCC CGTTTGAATT TAGGATCAAT AATAATTTGC 12720  
 TCAATTTTCAG CATTTAGTTC AATATTAACG CCTAAGTCTT TATTTAATTG CGCTAGCCCT 12780  
 TGAGCCATGC CATACATACC GCCTTTAATA AAATGCACAC CAAACATCAT TTCAATCATA 12840  
 GGAATAATTG AATATAGTGA CGGGCCTCGT TTTGGATCAA TTCCTATGTA TAACGTTTGA 12900  
 10 AACGCTAAAA GCTTTTGTAT CTTTTCGTTA TCAATATAAT GTTCAATTAG CTGATCTGCA 12960  
 TGATTTAACG TTTTAACTT AGCACCTTGC ACAAGTGACG TCATATTATA AAAGTCACTC 13020  
 GGTTCGCGAT ACGTTCCTTC TAAGAAATAG CGACGTGCAA TTTCATATTT TTTATAAACA 13080  
 15 TCCGTTAAAA AGGACATAAA ACCATGCGTT GAACCAGGTT CTATACTTTC TAGCATTGTC 13140  
 TGTAATTCAG CTAAATCTGT AGGCACCGTT ATACGATCAT CGTGGTCAAA ATACACATCG 13200  
 TAAATATAAC GTAATTGTCT CAATTCAATA TAATCTTCAT AATTTTTACC ACACGCTGTA 13260  
 20 AAAACATCTT TATAAACATC TGGCATCATG ACAATTGTGG GACCCATATC AAATGTAAAG 13320  
 CCGTCTTTCT TTAATTGATT CATACGCCCG CCTACATTAT TATTTTTTTC AAATATCGTC 13380  
 ACTTCATGAC CTGAGAAGC AATACGGGCT GCCGCTGCTA ATCCTGTGAC ACCTGCACCA 13440  
 25 ATTACTGCAA TCTTCATTAT TCAACCACCT ATATTCTATG ATATTTACTA TTTATTTTCAT 13500  
 GAAACAACCT TGCCTTTTTC CTCTTATCCA CAAAAACACG TTCATGTAAT GTATAGTTAG 13560  
 CCTGTCTCAC TTCGTCCAGT ATTCAATAT ATATACGTGC TGCTAATTCT ATGATTGGTT 13620  
 30 GTGCTTCAAT ACTAAATACT TTGATTTGAT CCATAACATC TTGAAAATCT TTTCTGCGA 13680  
 TAGCTGCATA ATATCCCAT AAGTCAATAT AATGATTATT AACACCATT TGGTACACTT 13740  
 35 CAGCAATATC AACTTCATAT TGCTTTAATC GTTGCTTACT AAAATATATC CGTTCATTGT 13800  
 CAAAATCTTC ACCGACATCT CTTAATATAT TAAGGGGATC CTCTAGAGTC GACCTG 13856

## (2) INFORMATION FOR SEQ ID NO: 32:

- 40 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 10088 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 45 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

50 ATATATAAAT ATAGATTAAG TATATAGATT AATCAACTTT TTTGGAAGAG CAAATCACGC 60  
 AATCAACAAA TAATATAAGA AGTTTTTGCG ATAGTTTTAA AATAGCTGTA ATAGAATACT 120  
 AAATGTGACA AACTTAGAAC TAATATCAAG TGTGATGTT TTGAATATAA AAATGCTAAT 180

	ATAATTGGTT AATATATGAG TAATTAGAAA ATAGACAAAG GATGACGATT TATGTATATC	300
	AATATGAAAG ATTATGGGTT AACAGGCATA AACAAAACTA AAGATACTCG AGCAATACAA	360
5	CGTGCGTTAA ATCGTGGAAG ATGTAAACCA ACGACAGTTT ATATACCGAA AGGGACGTAT	420
	GATATTTGCA AACCATTAAAC GATATATGGC AATACAACAC TTTTGTAGA TAATGAAACT	480
	ATTTTACGCC GATGTCAITC TGGTCCTTTA TTAAAAATG GTCGTCGCTT TGGTTTTTaT	540
10	CGTGTTTATA ATGGACACAG TCATATTCAT ATTAAAGGCG GCAAGTTTGA TATGAATGGT	600
	GTATCGTATC CTTATAACAA TACAGCTATG TGCATTGGGC ATGCTGAAGA TATTCAATTA	660
15	ATAGGTGTGA CCATTAAGAA TGTAAGTAGT GGTTCATGCA TTGATGCTTG TGGGATTAAC	720
	GGACTCTATA TTAAAAAGCTG TTCATTTGAA GGATTCATAG ACTATAGTGG CGAACcTTTT	780
	ATTCTGAAGC AATACAATTA GACATTCAAG TACCTGGTGC TTTTCCAAAA TCGGAACgA	840
20	CAGATGGTAC GATAACGAAA AATGTCATTA TCGAAGATTG TTATTTTGGG CCTTCAGAAT	900
	TGCCCCGAAAT GGGAAAGTTGG AATCGTGCTA TTGGCTCACA TGCAAGTAGA CATAATCGAT	960
	ACTATGAGAA TATTCATATT AGAAATAATA TATTTGAAGA TATACAAGGT TATGCATTAA	1020
25	CTCCCTTGaA GTATAAAGAT GCTTTCATTA TTAATAATAA GTTTATTAAC TGTGaGGGTG	1080
	GCATTAGATA TTTAGGAGTT AGAGATGGTA AAAATGCAGC AGATGTGaTG ACAGGaAAAG	1140
	ACTTAGGTTT CCAAGCAGGC ATAAATATGA ATATAATTGG AAATGAATTT AAAGGATCAA	1200
30	TGTCTAAAGA TCGGATACAT GTACGTAATT ATAATAATGT TAAACATAAA GATGTATTAA	1260
	TCGTTGGGAA TACATTCAAT AATTCGACTC AATCAATTCA TTTAGAAGAT ATTGATACAG	1320
35	TGTTTTTAAG TCCTGTTGAA GCGGGTATTC AAGTTACTAC AATCAATGTA GATGAAATAA	1380
	AAAAGTAAAA AGTTTCGCAT GACATTAGGA TTAAGAATAG TAGATAATTT TTGAAAGCGC	1440
	ATTēATAAAA CGGTATAAAT ATGCTATAAT AAACCCAATT ATCTGATAAA AGGGGTATTT	1500
40	TGACGGTAAT GATAATACAA GATAGACAAC TTTCTATACT CTAATATAGT GAGTTGAAGT	1560
	AGCTTGTCAT AATCATCATG AGGGGGAAAT TTATGGCTTA TTTCAATCAA CATCAATCAA	1620
45	TGATATCGAA AAGGTATTTA ACATTCTTTT CAAAATCAAA GAAAAAGAAA CCGTTTAGTG	1680
	CGGGACAACT TATTGGACTA ATATTAGGTC CATTACTTTT CCTATTAACA TTATTATTCT	1740
	TTCATCCACA AGACTTACCT TGGAAAGGCG TCTATGTTTT AGCGATTACT TTATGGATTG	1800
50	CGACTTGGTG GATTACTGAA GCAATTCCTA TTGCAGCAAC GAGCTTATTA CCAATTGTGT	1860
	TATTACCATT AGGTCATATA CTTACACCAG AACAAGTATC ATCCGAATAT GGCAATGATA	1920
55	TTATCTTTTT GTTTTTAGGT GGATTTATTT TGGCAATTGC AATGGAAAGA TGGAATTTAC	1980

	TTGGATTTCAT GGTGGCAACA GGATTCTTAT CTATGTTTGT ATCGAACACT GCAGCTGTAA	2100
	TGATTATGAT TCCGATTGGT TTAGCAATTA TTAAGGAAGC ACATGATTTA CAAGAAGCCA	2160
5	ATACGAATCA AACAAGTATT CAAAAGTTTG AAAAATCTCT AGTTTTAGCA ATTGGCTATG	2220
	CAGGTACGAT TGGTGGCTTG GGTACATTAA TCGGAACCCC GCCATTAAAT ATTTTAAAAG	2280
	GACAATACAT GCAACATTTT GGACATGAAA TTAGTTTTGC TAAATGGATG ATTGTAGGGA	2340
10	TTCCAACGGT CATTTGTTTTG TTAGGTATTA CTTGGCTCTA TTTAAGATAT GTTGCCTTTA	2400
	GACATGATTT GAAATATTTa CCTGGTGGTC AGACGTTAAT TAAACAAAAG TTAGACGAGC	2460
15	TTGGCAAAAT GAAGTATGAA GAAAAGGTAG TACAACTAT CTTTGTACTT GCTAGCTTAT	2520
	TATGGATTAC AAGAGAGTTT CTTCTGAAAA AATGGGAAGT TACGTCATCT GTTGCAGATG	2580
	GTACGATTGC TATTTTTATA TCAATATTAT TATTTATTAT TCCAGCTAAA AATACTGAAA	2640
20	AACATCGCCG TATCATTGAC TGGGAAGTTG CAAAAGAGCT CCCTTGGGGT GTATTAATTT	2700
	TATTTGGTGG CGGTTTAGCA TTAGCGAAAG GTATTTCTGA AAGTGGTTTA GCAAAATGGT	2760
	TAGGCGAACA GTTGAAATCA TTAAATGGTG TTAGTCCGAT TCTTATTGTA ATTGTCATAA	2820
25	CAATCTTTGT CTTATTTTTA ACTGAAGTGA CATCTAATAC TGCAACTGCA ACGATGATTT	2880
	TACCGATTTT AGCAACGTTG TCTGTTGCTG TTGGAGTGCA TCCATTACTA CTTATGGCAC	2940
	CTGCAGCTAT GCGGGCTAAC TGTGCATACA TGTTACCAGT AGGGACACCA CCGAATGCAA	3000
30	TTATCTTTGG TTCTGGTAAA ATATCTATCA AACAAATGGC ATCAGTAGGA TTCTGGGTAA	3060
	ACTTAATCAG TGCAATAATT ATTATTTTAG TCGTGTATTA TGTAATGCCT ATAGTTTTAG	3120
35	GTATTGATAT AAATCAACCA CTGCCATTGA AATAGTAATT GCAGATTAGA ACGAAAAATA	3180
	AAAGGTTACA TTAGCAATTG CTTGGACGAG TGGTAACGAA ACGTATACCG CAGCATCGTG	3240
	TAA <del>A</del> AACAAT ACAACAAAA GAAAGTCAAC CAAGGATGGA TTCCTATTTT AATCCTTGGT	3300
40	TGACTCTTTA TTTTATTTAA ATTGTAGAAC CTAGAAAATA AAGTTTAATT AAAAGCACCA	3360
	ATCATTTCTA CTTTGAAATC TAAGGTTTCT AAAATAGCAA TGACTTTCTT TATATCGGTT	3420
	GTAATTGCAG AATCAGCCTG AACGAAAAAT CGATACATAC CTAATTGTGT TTTTAAAGGA	3480
45	CGAGACTCAA TCCAGGATAA ATTAATATTA AACAAAGCAA ATGTATTAAG CACACTTGCT	3540
	AACAACCCAG GTTTATCATG CATTGGTGTA ATTA <del>A</del> AACA TCAATGATGT CGCATTTTGA	3600
50	TCAAATTGCT GCTGATTTT TATAACTAAA AAACGTGTCA CGTTATGTGG ATAGTCTTCA	3660
	ATATGTGTAT CAATAGGTGT AAAACCATAA GctTCGCCAC TACCTAAAGG TGCAATTGCT	3720
55	GCAACGCCAT TTTCAATTTT AGTCAA <del>A</del> CTT TGAATTGTAC TGTGACATA ATCATAGTCA	3780

	TTTTTAATAT CAGAAATGGA ATCTGTTCCA TTACCATATA ATGCAAAGTT AATATCTAAA	3900
	CGTATTTTCAC CGTGTGCAAA GACATCTTGC TGTGCAAGTG CATCTGCCAC AATGTTGATT	3960
5	GTTCTTTCTA TAGAATTTTC AATAGGGACA ACACCAATCG ATGTGTCATC ATCTGCAACT	4020
	GCCTTGATGA CTTCAAATAA ATTTGACTTT GGTGAAAAG TTGCTTCATT TTCAGAAAAA	4080
10	TACTGACGAC AAGCCAAATA TGAAAATGTA CCTTTAGGGC CTAAATAATA TAATTGCATA	4140
	TGCTACACCT CTACTAACTT AATGATGGAA AGGGCACTGG TTAGCATTTG ATTCTTTCTT	4200
	TTTATAGAAA AAGTTTGGAT CTTTTACTGT ATTGTCATAT CCGTGATGAT AATTTGACGT	4260
15	CAATGTTGGA GATAATGGCG GTGCTAGCCA AGACCATTTT CCGGTAAGTT GACGACCTTG	4320
	TTGTGCTTCG TTACGTTCTGA ATAGTTCTGAA TTGCTTTGCA GCGGTCAAAT GATCGACAAT	4380
	TGATACGCCT TCTTTTTTAA AGGAATGATA CACAGCATAG TTCAATTCAA CAAGTGCTCG	4440
20	ATCTTTATTA AATGAATTAT TTTTAAGTGT ATCAAATTCA AACGCATCTG CAACTTTTTTC	4500
	TAGTAAATTG TAACGGTAAT CATCAATAAA GTTACGTACG CCAATTTTCAG TTACCATATA	4560
	CCAACCGTTA AAGGGTGCAG TTGGATATAC AATGCCACCG ATTTTTTAAGT CCATATTGGA	4620
25	AATGATAGGG ACTGCATACC ATTTTAAGTT CAATTTTCTT AATTTTGGAT AATGATTATG	4680
	TTCAATAGGT ACTTCTTTAA TTAATGAAGT AGGATATTCG TAAAATTTAA CTGACTCATT	4740
30	AGGTAATTGG TAAATCAGTG GTAACACGTC AAAATTAGTA CCTTTTCCTT TCCAACCTAA	4800
	GTGATTTGCT AAGCGTGTA CTTCTTTTTC AGCAGGATCA CCACAATTGT CATAGCCAGC	4860
	ATAGCGAATT AATTGATTGT TGAAAATTTT AGGTCCATCC TTTGGAGCAT ATATAGTAAT	4920
35	ATACGGCTTT AATTTACCTT CATTTGTAGC CTGTGTAATA TGATAAGTAA TTGATGATAA	4980
	GAACGATGCT TCGTCAGTAA CATCTCTTGC ATCAATGACA TTTAACGAAT CCCAAAATAA	5040
	ACGA <sup>1</sup> CCAATG CAACGATTG AATTACGCCA AGCCATTTTA GCACCATAAA TAAGTTCTTC	5100
40	TTCTGTATGT GTATATGTCC CAGTTTCTTT TATTTCTAGT TCAATGTCAT GTAAACGTTT	5160
	ATTGATAATT TCGGTTTCAT AATGACACTC TTTATACATG TTTTCTATGA AAGCTTGAGC	5220
45	CTCTTTAAAT AACATTAACA ACACCTCGCT TTATATTATA GTCTACATTA TTAAAATACT	5280
	CTTAAAAATT ATGTATATGT CATTAAATTG TTGGTTGATT TTAATTAAAA GTATGGAAAT	5340
	TAAGGGGCTC TTATGTATAT AAAAAAATGA ATTATGATAA AATGTAAGAA AATATTTAGG	5400
50	TCGATTGGAG AGATACAAGT GTACCAATTA GAAGACGACA GTTTAATGTT ACATAATGAC	5460
	TTATATCAAA TAAATATGGC TGAAAGTTAT TGGAATGATA ATATTCATGA AAAAATGGCT	5520
55	GTATTTGATT TGTATTTTAG AAAAATGCCA TTTAATAGTG GCTATGCTGT TTTTAATGGT	5580

	TTAAAGTCTA TTGGCTACAA GGATGATTTT TTATCATATT TAAAAGATTT AAAATTCACA	5700
	GGCAGCATCC GTTCGATGCA AGAAGGCGAA TTATGCTTTG GTAACGAACC ATTGTTACGC	5760
5	GTAGAAGCAC CATTGATTCA AGCGCAATTA ATAGAAACAA TTTTATTAAA CATTGTAAAT	5820
	TTCCATACAT TAATTACAAC AAAGGCTAGC AGAATTCGTC AAATTGCATC AAATGATAAA	5880
10	TTAATGGAGT TTGGTACACG TCGTGCACAA GAAATTGATG CAGCATTGTG GGGCGCTAGA	5940
	GCTGCTTACA TCGGGGGCTT TGATTCTACA AGTAATGTTA GGGCGGGGAA ATTATTTGGT	6000
	ATACCTGTGT CTGGTACACA TGCACATGCA TTTGTCCAAA CTTATGGAGA CGAATATGTT	6060
15	GCCTTCAAAA AATATGCTGA AAGACATAAA AATTGTGTGT TCCTAGTAGA TACATTCCAT	6120
	ACTTTAAAT CTGGCGTGCC AAATGCAATA AAAGTTGCAA AAGAATTAGG TGACAAAAT	6180
	AACTTTGTAG GTATTCGATT AGATTCTGGA GATATCGCTT ATTTATCTAA AGAGGCAAGA	6240
20	CGTATGCTTG ATGAAGCAGG ATTTACTGAA ACTAAAATTA TCGCGTCTAA TGATTTGGAT	6300
	GAAGAAACGA TTACGAGTTT GAAAGCACAA GGTGCAAAAG TAGATTCTTG GGGCGTTGGT	6360
25	ACAAAGCTGA TTACAGGATA CGATCAACCA GCATTAGGTG CAGTATATAA ACTTGTAGCT	6420
	ATTGAAAATG AAGATGGTTC ATATAGTGAT CGTATTAAAT TATCAAATAA CGCTGAAAAG	6480
	GTTACGACGC CAGGTAAGAA AAATGTATAT CGCATTATAA ACAAGAAAAC AGGTAAGGCA	6540
30	GAAGGCGATT ATATTACTTT GGAAAATGAA AATCCATACG ATGAACAACC TTTAAAATTA	6600
	TTCCATCCAG TGCATACTTA TAAAATGAAA TTTATAAAAT CTTTCGAAGC CATTGATTTG	6660
	CATCATAATA TTTATGAAAA TGGTAAATTA GTATATCAAA TGCCAACAGA AGATGAATCA	6720
35	CGTGAATATT TAGCACTAGG ATTACAATCT ATTTGGGATG AAAATAAGCG TTTCTGAAT	6780
	CCACAAGAAT ATCCAGTCGA TTTAAGCAAG GCATGTTGGG ATAATAAACA TAAACGTATT	6840
	TTTGAAGTTG CGGAACACGT TAAGGAGATG GAAGAAGATA ATGAGTAAAT TACAAGACGT	6900
40	TATTGTACAA GAAATGAAAG TGAAAAAGCG TATCGATAGT GCTGAAGAAA TTATGGAATT	6960
	AAAGCAATTT ATAAAAAATT ATGTACAATC ACATTCAATT ATAAATCTT TAGTGTTAGG	7020
45	TATTTTCAGGA GGACAGGATT CTACATTAGT TGGAAAATA GTACAAATGT CTGTTAACGA	7080
	ATTACGTGAA GAAGGCATTG ATTGTACGTT TATTGCAGTT AAATTACCTT ATGGAGTTCA	7140
	AAAAGATGCT GATGAAGTTG AGCAAGCTTT GCGATTCATT GAACCAGATG AAATAGTAAC	7200
50	AGTCAATATT AAGCCTGCAG TTGATCAAAG TGTGCAATCA TTAAAAGAAG CCGGTATTGT	7260
	TCTTACAGAT TTCCAAAAAG GAAATGAAAA AGCGCGTGAA CGTATGAAAG TACAATTTTC	7320
55	AATTGCTTCA AACCGACAAG GTATTGTAGT AGGAACAGAT CATTGAGCTG AAAATATAAC	7380

	TAAACGACAA GGTGTCAT TATTAGCGTA TCTTGGTGCG CCAAAGGAAT TATATGAAAA	7500
	AACGCCAACT GCTGATTTAG AAGATGATAA ACCACAGCTT CCAGATGAAG ATGCATTAGG	7560
5	TGTAACCTAT GAGGCGATTG ATAATTATTT AGAAGGTAAG CCAGTTACGC CAGAAGAACA	7620
	AAAAGTAATT GAAAATCATT ATATACGAAA TGCACACAAA CGTGAACCTG CATATACAAG	7680
10	ATACACGTGG CCAAATCCT AATTTAATTT TTTCTTCTAA CGTGTGACTT AAATTAAATA	7740
	TGAGTTAGAA TTAATAACAT TAAACCACAT TCAGCTAGAC TACTTCAGTG TATAAATTGA	7800
	AAGTGTATGA ACTAAAGTAA GTATGTTTAT TTGAGAATAA ATTTTTATTT ATGACAAATT	7860
15	CGCTATTTAT TTATGAGAGT TTTCGTTACTA TATTATATTA ATATGCATTC ATTAAGGTTA	7920
	GGTTGAAGCA GTTTGGTATT TAAAGTGTA TTTGAAAGAGA GTGGGGCGCC TTATGTCATT	7980
	CGTAACAGAA AATCCATGGT TAATGGTACT AACTATATTT ATCATTAACG TTTGTTATGT	8040
20	AACGTTTTTA ACGATGCGAA CAATTTTAAC GTTGAAAGGT TATCGTTATA TTGCTGCATC	8100
	AGTTAGTTTT TTAGAAGTAT TAGTTTATAT CGTTGGTTTA GGTTTGGTTA TGTCTAATTT	8160
	AGACCATATT CAAAATATTA TTGCCTACGC ATTTGGTTTT TCAATAGGTA TCATTGTTGG	8220
25	TATGAAAATA GAAGAAAAAC TGGCATTAGG TTATACAGTT GTAAATGTAA CTTCAGCAGA	8280
	ATATGAGTTA GATTTACCGA ATGAACTTCG AAATTTAGGA TATGGCGTTA CGCACTATGC	8340
30	TGCGTTTGGT AGAGATGGTA GTCGTATGGT GATGCAAATT TTAACACCAA GAAAATATGA	8400
	ACGTAAATTG ATGGATACGA TAAAAAATTT AGATCCGAAA GCATTTATCA TTGCGTATGA	8460
	ACCTCGAAAC ATACATGGTG GATTCTGGAC TAAAGGCATT CGTCGTAGAA AGCTTAAAGA	8520
35	TTATGAACCA GAAGAACTGG AAaGTGTAGT AGAaCATGAA aTTCmAAGTA AaTGAGAAaTG	8580
	AAmCAATtGC TGATTGTTTG TCACGAATGA AAtGCAAGGG TATATGCCGG TAAAACGTAT	8640
	TGAATAAACC GTGTTTCAAG AGCAAAAAGA TGGCACGGTT GAAGTATCAC ATCAAGAAAT	8700
40	CGTTTTTGTA GGTAAAGAAA TCCAATAACA TAATCCAATT TAAATAAAGA CTATTTGAAG	8760
	AGGAAAGGCT ATTCAAAGTT TGAGTAATTT TACTTTGAAT AGCCTATTTG TTTATACATG	8820
45	CAAGATGCTC GATCCATATT GTATGAGAAA CCCCAGCAA GCTATATAAA GCATATGCTG	8880
	GGGGTTCTTA ATATTTTAAA AATTATTGTT AGATTATATA TATCGTCGCT TTTTCTAAAA	8940
	CAATCTCATC GCATGAAATT TTTTCTTCCT AGAGACCTTT AATAAGATTA ATAGTTTACT	9000
50	TAATCATATC TAGATAGTCT TATGACTTAT GCTTAATGAA AGTCATTCTA GGAGAAGTTC	9060
	CCAAAGCTTC TGTGTTTATA ATTGTTAGTA GTATTTTATT ATCATTTGGT ATAAATATTT	9120
55	CAATAACAAT TGAGCTATTA TTTTATTAT ATAATGTGAG TTGTTTGTGT TCTGTATTTA	9180

CATTTAATC TTGAGGATGC CATTCTCCCT CAATAATATT AAGATAATAC TTAGCCTCTG 9300  
 AATTACATTT GAATTTATCA ATACTAAATA ATTCAATTTG TTCCATAATA TTATTTACCT 9360  
 5 TTCTAAAATA CAAATTTTAA TAACCATAAA TAGATGAATA CCATCGATAA TGGTCGCCAT 9420  
 TGGATACTGG AATAACATTG TTTTGTAGCAT CTTGAGTCAT AAAACCATTG TCCCATGGAT 9480  
 TCCATATAAT TATAACCTCT TGTCCATTAT CTAATTTAGC GTTCCCAACA ACTGCCATGG 9540  
 10 CATGCCCTGC GTGCATACCA TTTCTTGATT CTACTCTACT ACCTAAAACA GCAATTCCTT 9600  
 TATTATTTTT AGTAAGATTG TCAACTTCAT TATATGTAGT CATTCTATTA AGAAGTTGTG 9660  
 15 GACTTCTTCC CTGAGTTTGT CCAAAATAAA TCATCTCTCT TGGCGTTAAA CCAGTAAATT 9720  
 GGAATCGTTG TCCTTGTAAG TTTGGGTGTA AAAATCTCAT CACAGCTTCT GCATGATATT 9780  
 TGTTAGTATT ATAAGTCGCA TTTAGTAATT CAGACATCGT ATAGCCTGCA CACCAACCAT 9840  
 20 TGTTACCTTG AGTTTCTCTT ATCTTGAAAT TCTCAAGTTT ATTTATATAT TGStCGTTGT 9900  
 AAGTATAATT ATTACTTTTA AATTGACTAG TTGGCATAGT GACAGAAGCT TTTTGCTTTA 9960  
 GTTGC GTTAC ATTATTGCCA GTAGGTATAC TCTCAGTCTT TnTnAACTnT nTATCTTCTA 10020  
 25 GACGTGGTGT TTTTAGTACT AGTTTAGCTT TATGATTTTG AGTACCACAT AGTAACCTTT 10080  
 TGAGTTGT 10088

## (2) INFORMATION FOR SEQ ID NO: 33:

- 30 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 7563 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 35 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

40 CGGAAACGnA CCCnATGCGT ATGCTTGACG TGCCAAAATT AAATACGAAG TTCATAGCTT 60  
 TGAGGTACCA GAAGAACATT TATCTGGTCA AGAAGTCGCA GnACTCATAC AAGCAAATGT 120  
 TAAAACAGTA TTTAAAACGC TTGTTCTAGA AAATACAAAA CATGAACATT TTGTATTTGT 180  
 45 TATCCCAGTA AGTGAAACTT TAGATATGAA AAAGGCAGCT GCTTTGGTTG GAGAGAAGAA 240  
 ATTGCAGCTT ATGCCTTTAG ATAATTTGAA AAATGTAACG GGATACATTC GTGGTGGGTG 300  
 TTCGCCTGTT GGTATGAAAA CATTGTTTCC AACAGTCGTT GACAAATCGT GTGAAAATTA 360  
 50 TAGTCATATC AGTGTGAGTG GTGGGCTTCG AACAAATGCAA ATCACAATAG CTGTTGAGGA 420  
 TTTGATTACA ATAATAAAG GCAAAATTGG AGCAGTTATC CATGAATGAT TAATAACAAC 480

55

	TGCCCACTC	CTTTTGATT	GAATTAGCAT	TTTACGATCA	TAAACAGTCA	TTATAATTGA	600
	GTATTTGAAC	ATAAAAATGT	AATTTTATCG	TAACAATTTG	AGTGTTTG	ATTGTTTTTG	660
5	GTAATTTATG	ATTGAAAAGT	GAAAGCGTAC	TCATTATAAT	ACAAAGTGAG	ATGGGGTGAT	720
	GATGATAATT	ACTGaAAAAA	GACACGAGTT	AATATTAGAA	GAACCTTCGC	ACAAAGATTT	780
10	TTTGACTTTA	CAAGAATTAA	TAGATCGAAC	TGTTGTCAGT	GCTTCAACAA	TACGArGAGA	840
	TTTATCTAAA	CTACAACAAT	TAGGGAAATT	GCAACGTGTG	CATGGTGGTG	CAATGTTAAA	900
	AGAAAATCGT	ATGGTTGAGG	CGAATTTAAC	TGAAAAATTA	GCAACGAATC	TTGATGAAAA	960
15	GAAAAATGATT	GCTAAAATAG	CAGCTAATCA	AATCAACGAT	AATGAATGCT	TATTTATCGA	1020
	TGCTGGTTCA	TCTACATTGG	AGCTAATTAA	ATATATTCAA	GCGAAAGATA	TCATTGTGGT	1080
	AACCAATGGT	TTAACACATG	TAGAAGCTTT	ACTTAAAAAA	GGTATTAAAA	CAATTATGCT	1140
20	AGGTGGTCAA	GTTAAAGAAA	ATACACTTGC	TACGATTGGT	TCTAGTGCTA	TGGAGATATT	1200
	AAGACGATAT	TGTTTCGATA	AAGCTTTTAT	CGGGATGAAT	GGATTAGATA	TTGAACTTGG	1260
	ATTAACACT	CCCCGATGAGC	AAGAGGCATT	AGTTAAACAA	ACAGCAATGT	CATTAGCCAA	1320
25	TCAATCATTT	GTACTTATAG	ATCATTCTAA	GTTTAATAAA	GTATATTTTG	CTCGTGTACC	1380
	TTTGCTAGAA	AGTACGACAA	TCATCACATC	TGAAAAAGCA	TTAAATCAAG	AATCGTTAAA	1440
	AGAATACCAA	CAAAAGTATC	ACTTTATAGG	AGGGACTTTA	TGATTTATAC	AGTGACTTTC	1500
30	AATCCTTCAA	TTGACTATGT	CATTTTACG	AATGATTTTA	AAATTGATGG	TTTGAACAGA	1560
	GCAACAGCAA	CATATAAATT	CGCTGGGGGG	AAAGGTATTA	ATGTCTCGCG	CGTCTTAAAG	1620
35	ACATTGGATG	TTGAGTCAAC	TGCCTTGGA	TTTGCAGGTG	GATTTCTTGG	GAAATTCAAT	1680
	ATAGATACAT	TAAATAACAG	TGCAATTCAA	TCGAATTTTA	TTGAAGTTGA	TGAAGATACA	1740
	CGTATTAATG	TGAAATTAAA	AACAGGACAA	GAAACAGAAA	TCAATGCACC	GGTCTCTCAT	1800
40	ATAACGTCAA	CACAATTTGA	ACAACTGTTA	CAACAAATTA	AAAATACAAC	AAGCGAAGAT	1860
	ATAGTTATTG	TTGCTGGAAG	TGTACCAAGT	AGTATTCCAA	GCGATGCGTA	TGCGCAAATT	1920
	GCACAAATTA	CAGCACAGAC	AGGTGCTAAA	TTAGTAGTCG	ACGCTGAAAA	AGAATTGGCT	1980
45	GAAAGCGTTT	TACCATATCA	TCCACTATTT	ATTAAACCTA	ATAAAGATGA	ATTAGAAGTG	2040
	ATGTTTAATA	CAACAGTGAA	CTCAGACACA	GATGTTATTA	AATATGGTCG	TTTGTTAGTT	2100
50	GATAAAGGTG	CGCAATCTGT	TATTGTCTCG	CTTGGCGGTG	ATGGTGCTAT	TTATATTGAT	2160
	AAAGAAATCA	GTATTAAAGC	AGTTAATCCA	CAAGGGAAAG	TGGTTAATAC	AGTTGGCTCT	2220
55	GGTGATAGTA	CAGTTGCAGG	CATGGTGGCT	GGAATTGCTT	CAGGTTTAAC	GATTGAAAAA	2280



	CGGGACGCTA TAGAAAAAAT AAAATCACAA GTTACGATTA GCGTACTTGA TGGGGAGTGA	2400
	AAATAATGAG AGTAACAGAG TTATTAACAA AAGATACAAT AGCAATGGAT TTAATGGCAA	2460
5	ATGACAAAAA TGGTGTTATT GATGAGTTAG TAAATCAATT AGACAAAGCA GGTAAATTAA	2520
	GTGATGTCGC GTCATTTAAG GAAGCGATTG ACAATCGAGA ATCACAAAGT ACAACTGGTA	2580
	TCGGCGAAGG TATTGCCATT CCACATGCCA AAGTGGCCGC AGTTAAGTCA CCAGCTATTG	2640
10	CGTTTGGTAA ATCTAAAGCA GGCCTAGATT ATCAAAGTTT GGATATGCAA CCAGCACACT	2700
	TATTCTTTAT GATTGcAGcG CCAGAAGGTG GCGCCCAAAC ACATCTAGAT GCTTTAGCTA	2760
	AGTTGTCTGG TATTTTAATG GATGAAAATG TACGTGAGAA ATTATTACAT GCTTCATCAC	2820
15	CTGAAGAAGT ACTAGCGATC ATAGATGAGG CTGATGATGA AGTGACAAAA GAAGAAGAGG	2880
	CAGAAGCTGA AGCACAAACA GTTGCAACTG CAGAACAATC ATCTAAACAA TCTAATGAGC	2940
20	CATATGTGTT AGCAGTAACT GCTTGTCCAA CAGGTATTGC ACACACATAT ATGGCACGTG	3000
	ATGCATTGAA AAAGCAAGCG GATAAAATGG GTATTAAAAT TAAAGTAGAA ACGAATGGTT	3060
	CAAGCGGCAT TAAAAACCAT TTAAGTGAAC AAGATATTGA AAATGCAACA GGTATCATTG	3120
25	TTGCTGCTGA TGTTTCATGTT GAGACGGATC GCTTCGATGG TAAAAATGTC GTAGAAGTAC	3180
	CAGTAGCAGA TGGTATTAAA CGCCCAGAAG AATTAATTAA TAAAGCATTG GATACAAGTC	3240
	GTAAACCTTT TGTTGCCCGT GATGGTCAAA GAAAAGGTAA CTCAAATGAC AGTCAAGAAA	3300
30	AATTAAGCCC AGGTAAAGCA TTCTATAAAC ACTTAATGAA CGGTGTTTCT AACATGTTGC	3360
	CACTTGTAAT ATCTGGTGGT ATTTTAATGG CAATTGTATT TTTATTTGGA GCAAATTCAT	3420
	TTAATCCAAA AAGCTCAGAG TACAATGCGT TTGCAGAGCA GCTTTGGAAC ATTGGTAGTA	3480
35	AAAGTGCATT CGCGTTAATC ATTCCAATTT TATCTGGATT CATTGCACGT AGTATTGCGG	3540
	ATAAAcCTGG TTTCGCTTCA GGTCTTGTAG GTGGTATGTT AGCAATTTCA GGTGGTTCAG	3600
40	GATTTATTGG TGGTATTATT GCAGGTTTCT TAGCAGGTTA CTTAACACAA GGTGTTAAAG	3660
	CCATGACACG TAAGTTACCA CAAGCATTAG AGGGATTAAA GCCAACATTA ATTTATCCAC	3720
	TATTAACAGT GACGGCTACA GGCTTATTGA TGATTTATGC CTTTAATCCA CCAGCATCTT	3780
45	GGTTAAATCA TTTGTTATTA GATGGATTAA ACAATTTATC AGGTTCTAAT ATTGTATTAT	3840
	TAGGTTTAGT TATTGGCGCT ATGATGGCGA TTGATATGGG CGGTCCATTG AACAAAGCGG	3900
	CATATGTTTT TGCAACAGGT GCGTTGATTG AAGGTAATGC AGCACCAATT ACAGCTGCAA	3960
50	TGATTGGTGG TATGATTCCA CCGTTAGCAA TTGCGACAGC GATGTTAATT TTTAGACGTA	4020
	AATTTACAAA AGAACAACGT GGTTCATTA TCCCTAACTA TGTGATGGGT ATGTCATTTA	4080
55		

	TGATTGGTTC AGGTATAGGT GCGCAATTG CTTTAGGCTT AGGTTACCGA ATTACTGCGC	4200
	CACATGGTGG TATTATTGTA ATTGTTGGTA CTGATGGTGC ACACTTACTT CAAACTCTTA	4260
5	TTGCACTTCT AGTTGGCACA TTAGTTTCAG CATTAAITTA CGGTTTAATC AAACCAAAGT	4320
	TAACTGAAAC AGAAATCGAA GCTTCAAAAT CAATGGACGA GTAGTTTAA TGATGTAAAA	4380
	TGATTGTTAG CAAAGAGCTT CATATTAAGT TGTATGTTCA ATGAATATAT GTTAGTTTTA	4440
10	TATATCGTGT TAACGGTAGC TTATACAAAG CTGTAAAAAC ACTTCTATT AATTCAGTTT	4500
	TTATGAATTG ATATGAAAGT GTTTTTATTT TTAGATAAAT GAATGAAGAA ATAGACACCA	4560
15	CAAATGTATA GACTTTTTTA ATATTTTGCA AAAAGTTATG CCAAACGAAG CAGATATAGT	4620
	AAAATATGAG TGTCTTAAAG TGAAAATTTA TAAATAAAGA AGGGTTTATA CGTGTCAGAA	4680
	TTAATTATAT ATAACGGCAA AGTTTATACT GAAGATGGCA AAATCGATAA TGGTTACATT	4740
20	CATGTGAAAG ATGGACAGAT TGTGCAATT GGAGAAGTGG ATGATAAAGC AGCAATTGAT	4800
	AATGATACGA CAAATAAAAT TCAAGTGATT GATGCTAAAG GTCATCATGT ATTACCAGGT	4860
	TTTATTGATA TACATATTCA TGGTGGTTAT GGTCAAGATG CAATGGATGG GTCATACGAT	4920
25	GGCTTAAAT ATCTATCCGA AAATTGTGTG TCTGAAGGGA CGACATCATA CTTGGCCACT	4980
	ACAATGACGC AATCGACTGA TAAAATAGAT AATGCACTTA CAAATATTGC TAAATATGAA	5040
	GCGGA <sub>g</sub> CAAG ATGTTCAAA TGCAGCGGAA ATTGTAGGTA TACATTTAGA AGGACCATTT	5100
30	ATATCTGAAA ATAAAGTTGG TGCTCAACAT CCGCAATACG TTGTACGCCC ATTTATCGAT	5160
	AAAATTAAAC ATTTTCAAGA GACTGCTAAC GGATTAATAA AGATTATGAC GTTTCACCT	5220
35	GAAATGAAG GTGCAAAAGA AGCGCTTGAA ACGTATAAAG ATGACATTAT TTTTCAATT	5280
	GGTCATACAG TAGCAACATA CGAAGAAGCA GTTGAAGCTG TTGAGCGAGG AGCTAAACAT	5340
	GTCACGCATT TATATAATGC AGCGACGCCA TTCCAACATA GAGAACCAGG TGTTTTTGA	5400
40	GCAGCATGGT TGAATGATGC TCTACATACC GAAATGATTG TTGATGGCAC TCATTCTCAT	5460
	CCGGCATCGG TTGCAATTGC TTACCGTATG AAAGGTAATG AACGTTTTTA TTTAATTACC	5520
	GATGCAATGC GTGCAAAAGG TATGCCTGAA GGAGAATATG ATTTGGGTGG ACAAAAAGTA	5580
45	ACTGTTCAAT CGCAACAAGC ACGTCTTGCA AATGGTGCGC TTGCTGGTAG TATTTTAAAA	5640
	ATGAATCATG GGTACGTAA CTTAATATCA TTTACAGGTG ATACATTAGA TCATTTATGG	5700
	CGAGTAACAA GTTTAAATCA AGCCATTGCA TTAGGTATCG ATGATAGAAA AGGTAGTATT	5760
50	AAAGTAAATA AGGATGCAGA TCTTGTTATT CTAGATGATG ATATGAATGT AAAATCTACA	5820
	ATAAAACAAG GCAAGGTTCA CACATTTAGC TAATAAATAA TCATAATTAA ATGTATGCAA	5880
55		

	TTTTCTGGGG GTGTCTAAAT GGAAGGCCA TAACATGTAG TTGTAATTTA AGTCATAGTG	6000
	ATAAATTTGA ATGCGTGTTA CCCATGAGTG ACACATATAA CATGGAGGTG AATCCCTAGA	6060
5	AATAGGGAAT TAATTGAAA CTTGACCAT AATTAGTTTG ATTATATTTA TTCTATTAAT	6120
	TGCATTAACC ACTGTATTTG TTGGTTCAGA ATTTGCATTA GTAAAAATTA GAGCAACAAG	6180
	AATTGAACAG CTAGCAGATG AAGGAAATAA ACCTGCTAAA ATAGTAAAAA AGATGATTGC	6240
10	TAATCTAGAT TAITATCTTT CTGCTTGTC AATTAGGTATA ACAGTAACAT CTTTAGGGTT	6300
	AGGTTGGCTT GGTGAACCAA CGTTTGAAAA GCTATTACAC CCAATATTTG AAGCAATCAA	6360
15	TTTACCAACT GCATTAACGA CGACGATTTT GTTTGCAGTG TCATTTATAA TCGTTACGTA	6420
	TTTGCATGTA GACTTGGTG AATTAGCGCC TAAATCTATA GCTATTCAAC ATACTGAAAA	6480
	GCTTGCTTTA GTATATGCAA GACCATTGTT CTATTTCCGT AACATTATGA AACCATTGAT	6540
20	TTGGCTGATG AATGGTTCTG CACGTGTTAT TATTAGAATG TTTGGTGTA ATCCTGATGC	6600
	CCAACTGAT GCAATGTCAG AAGAAGAAAT CAAAATTATT ATTAACAATA GTTATAATGG	6660
	TGGAGAAATC AACCAAACTG AATTGGCATA TATGCAAAAT ATCTTTTCAT TCGATGAAAG	6720
25	ACATGCAAAA GATATAATGG TACCTAGAAC TCAAATGATT AACTAAATG AACCTTTTAA	6780
	TGTAGACGAA TTACTAGAAA CAATAAAGA ACATCAATTT ACGCGTTATC CAATTACTGA	6840
	TGATGGTGAT AAAGACCACA TTAAAGGATT TATTAACGTC AAAGAATTTT TAACTGAATA	6900
30	CGCTTCTGGA AAAACGATTA AAATAGCAAA CTATATaCAT GAGTTGCCAA TGATTTTCAGA	6960
	GACAACACGT ATCAGTGATG CATTAATTAG AATGCAACGT GAACATGTAC ATATGAGTCT	7020
35	TATTATAGAT GAATATGGTG GAACGGCAGG TATTTTAAAC ATGGAAGATA TTTTAGAAGA	7080
	AATCGTTGGA GAAATTCGTG ATGAATTTGA TGATGATGAA GTGAATGATA TCGTTAAAT	7140
	TGATAATAAG ACATTCCAAG TAAATGGCAG AGTACTATTG GATGATTTAA CTGAAGAGTT	7200
40	CGGTATAGAA TTTGATGACT CTGAGGATAT TGATACGATA GGTGGATGGT TACAATCTCG	7260
	TAATACCAAT TTACAAAAAG ATGATTACGT GGATACAACT TATGATCGCT GGGTTGTTTC	7320
	AGAAATCGAT AACCACCAAA TTATTTGGGT GATATTAAAC TATGAATTTA ATGAAGCGAG	7380
45	ACCTACTATC GGACAGTCTG ATGAAGATGA AAAATCAGAA TAGATATTAA TATATAAACC	7440
	AACTAAGAAT GATTAAATTC ATTTTGGTT GGTTATTTT TTGACTAAAA TTAAnGAAAA	7500
	GTGAAAATAG TATTGGAAC CAATATCTTT AATGATTTAA TGAATAAnTT TTATTGAAAG	7560
50	CGA	7563

(2) INFORMATION FOR SEQ ID NO: 34:

55

(A) LENGTH: 3492 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

10	TTATATCAAC TTCATGGCGG AACCATTGAT GACCCATTAG ACGAAACAAT AAGCGCATTT	60
	sATGAATTGA AACAGAAGG AATTATACGT GCTTACGGTA TTTCTTCTAT TCGCCCAAAT	120
	GTAATTGATT ATTATTTAAA ACATAGTCAA ATCGAAACGA TAATGTCTCA ATTCAATTTG	180
15	ATTGATAATC GTCCAGAATC ATTATTAGAT GCAATTCACA ACAATGATGT TAAAGTATTG	240
	GCAAGAGGAC CTGTGTCTAA AGGATTATTA ACTTCAAACA GTGTTAATGT GCTCGACAAT	300
	AAATTTAAAG ATGGTATTTT TGATTATTCT CATGATGAAT TGGGTGAAAC AATAGCCTCT	360
20	ATTAAAGAAA TTGAAAGTAA TTTATCTGCA TTGACATTTA GTTATTTAAC ATCACATGAC	420
	GTGCTTGGTT CCATCATTGT AGGTGCAAGT AGCGTCGACC AATTAAAAGA AAATATTGAA	480
25	AACTATCATA CTAAAGTTAG TTTAGATCAG ATTAAAACAG CAAGAGCTCG TGTAAAGGAT	540
	TTGGAATATA CCAATCATTT AGTGTAGAAG TCATTTTCAG TAATAAAAAC AGCAGCATGA	600
	GGCGTTTCAT TATAAAAATG CCTTACTGCT GTTGTTTATG TACAATTCGC TATAATTTAT	660
30	GATTATGATT ACTCACTTAT GATAGAAATT AAAGCGTTGT CCTCACGCAT CAGTATTTAG	720
	TAATTTCGCC TTGCGGCATT GCCTTAAGCA AACTTCTGCC ACTTCATCTC TTAATAATTT	780
	TATTAACA TCTTTCTATA TTTCACITCG CATGTTGATT CATCATTATT AGTTATTATT	840
35	TGTACACCCA GCACATTTCC TTGCAACACA AGTAGTTTGA ATTTTTCACA AGTATAATAT	900
	AATGTACCGT CTGAAATTTG GTCTACAGAA ATATCGCCTA AAATATCCAG CACTGTAAAT	960
	TCTTCAAATA CTGATAGTTG TTCCGCATAT CGTACACAAA GTCTTACCAC ACTCTCCGAT	1020
40	TGACAGTTCA TTGCCATCCC ACCTATTTAT GCTTTATTTT TAAATAATTT AGGGAAACAT	1080
	CGTTCAAAAA ATCTAGGCGC AATTTGATAC ATTTTCAACG CATGATGCAT CCATTTAGGC	1140
45	CGATTAAITTT CCAATTGTTT TGTTTTAATG CCATAAATGA TATCTTCTGC AAGCTGATTA	1200
	GCATCAAGCA TAATTTCCCC CATCTTTTTA gCATACTTCA TTGATGGGTC GGCTTTTTGA	1260
	TGAAAAGGTG TATCAATCGG GCCAACATTA ACTGTCAATGA TATGTAAGTT TGGTGACTCT	1320
50	AGTCTTAAAG CATTCAATTA TGCATAAAAC CCTGCTTTCG ATGCCCCATA ATGTGCAGCA	1380
	TTTGCTTG TGGAATATGC AGCTTGACTT GAAATACCTA CAATATGTGC GTTAGATGTT	1440
55	AAATATGGTC TCAACACAGT ATATAAACA TTAATACTAA TTAAATTAAG CTGATACGTT	1500

	TAAATGAATC CATCGAATGA TGTATTGTCT TCAAATTGCA GTGCCTGTAT CGACTTCAAA	1620
	TCATTTAAGT CACAAGGAAT AACATTTATA GTTTTCCCCA ATTCCTGTTC AAAGATTCTA	1680
5	GTTGCTTTAT CAACATCACG CACCAACAAC GTTACATGCA CTTTATTTTC TAGTAACTTT	1740
	CGGACAATCG ATAAACCTAA ACCACTCGTA CCACCAGTCA CTATAAAATG TTGTCCTTTC	1800
10	ATCAATTAACT CTTCTTTTTC AATTATATAG AATGCAATTT ATCAACTTTA CATAATTGAG	1860
	ACAAGTTGAT TATCTTTCTT AATATATATA CAATAATAAG AAAATATAAC ATACAAATCA	1920
	AAAATAAAG GGATGTGACG TTAATGTAAC TCGTATTTTA TGGAGCTGGT AATATGGCAC	1980
15	AAGCTATATT TACAGGTATT ATTAACTCMA GCAACTTAGA TGCCAATGAT ATATATTTAA	2040
	CAAATAAATC TAATGAACAA GCTTTAAAAG CATTGCTGA AAAACTAGGT GTTAACTATA	2100
	GTTATGATGA TGCGACATTA TAAAAGATG CAGATATGT ATTTTATAGT ACCAAACCAC	2160
20	ATGACTTTGA TGCTCTAGCA ACACGCATCA AACCACATAT TACAAAAGWC AATTGCTTCA	2220
	TTTCAATTAT GGCAGGTATT CCGATTGATT ATATTAAACA ACAATTAGAA TGCCAAAATC	2280
	CaGTGTCTAG AATTATGCCA AACACAAATG CGCAAGTTGG ACACTCTGTT ACTGGCATT	2340
25	GTTTTTCAAA CAACTTTGAC CCTAAATCTA AAGATGAAAT TAACGATTTA GTTAAAGCAT	2400
	TTGGTTCTGT AATTGAAGTA TCAGAAGATC ATTTACATCA AGTAACAGCT ATCACCAGAA	2460
	GCGGCCCAGC ATTTTATATAT CATGTATTCG AGCAATATGT TAAAGCTGGT aCsAAACTTG	2520
30	GTCTAGAAAA AGAACAAGTT GAAGAATCTA TACGCAACCT TATTATAGGT ACAAGTAAGA	2580
	TGATTGAACG TTCAGATTG AGCATGGCTC AATTAAGAAA AAATATTACC TCTAAAGGTG	2640
35	GTACGACACA AGCTGGCCTT GATACATTGT CACAATATGA TTTAGTATCT ATTTTCGAAG	2700
	ATTGTCTAAA CGCTGCCGTC GACCGTAGTA TTGAACCTTC TAATATAGAA GACCAATAAA	2760
	AACAATACCCG CCAACACATG TATGCATCAT CGCAAGCACT GTGTTTGACG GGTTATTTTT	2820
40	ATAATTTATT GTTATTTGGC AAGCATTGTT TATTACTTTG TCATTAGATT TTAAAACTAT	2880
	CAAAATCTTT TACAAAATTA AAATTAGGTG TATCTTCATT TTGTATCAAT GTTTGATAAA	2940
	TTTCATTTAT ATCTTCTGTA TTATAGCGAT TGCTCAAATG TGTAATCAAC GTACGTTTAA	3000
45	CATTGGCTTC TTTTATCAAT GCAAATACGT CTTCAATATG GCTATGATGA TAATTGTTGG	3060
	CTAAATGCTT TTCACCATCT ATATAGGTCG CTTCAATGAC CATCACATCA GCATCTCTAG	3120
	AAATCACACG TTCATTAGAA CATGGTTTTG TATCACCAA AATTGCTACA ACTGGACCCT	3180
50	GTTTGGACTC ACCTCTAAAA TCTTTTGATT GATAAACTTG ACCATTATGT TCAAATGTAT	3240
	CATGAGATTT TACTTCTTGA TATTTAGGAC CTGGTTCAAG ACCAATGTTT TTAAACGCTT	3300
55		

CATGATTAAG TAAATGCGCC TCTACAGTAA AACCATCCAT GATGATATGT CAGATGATCA 3420  
 TCGATTTCAA TATATGtAAT TGGATAGTTT AAATGTGACT CTGATAAATT CATAGACATT 3480  
 5 TCCACATATG CT 3492

(2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1973 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

ATCTAGCGGT ACAAGCGTCT TGGAGGCTAG TATGTTGAAC ATTGTAAACC CTGAAGATCA 60  
 20 CTTGCGTTGTC ATTGTTTCAG GTGCCTTTGG TAACCGATTT AAACAAATTG CACAAACTTA 120  
 TTACAAAAT GTGCATATTT ATGACGTAAC ATGGGGAGAA GCTGTAGATG TCAAAGATTT 180  
 CATCAATTTT CTTTCAACTT TAAATGTTGA AGTTAAAGCA GTATTTAGTC AATATTGCGA 240  
 25 AACATCTACG ACAGTGCTAC ACCCTATTCA CGAGTTAGGA AATGCCATTA ATCAATTTAA 300  
 TAGTAATATT TATTTTGTAG TTGACGGCGT AAGTtGCATT GGTGCTGTTG ATGTTGACAT 360  
 TAACAAAGAT AAAATTGATG TACTTGTTTC TGGTAGTCAA AAAGCAATTA TGTTACCTCC 420  
 30 AGGATTAGCT TTTGTAGCTT ATAGCCACCG TGCAAAAGAA CATTTCAAAG AAGTAACTAC 480  
 GCCAAAATTT TATCTAGACT TAAATAAATA CATTTGCTCA CAAGCTGACA ATTCTACACC 540  
 GTTCACACCA AATGTGTCTT TATTTAGAGG TGTAAATGCA TACGTTGAAA CCGTAAAAGC 600  
 35 AGAAGGTTTC AATCACGTAA TAGCACGACA CTATGCAATT AGAAATGCAT TAAGAAGCGC 660  
 CTTAAAGCA TTAGATTTAA CTTTATTAGT CAATGATAAA GATGCATCTC CAACGGTTAC 720  
 40 AGCATTCAA CTAATACAA ATGATGAAGT GAAAATAATC mAAGATGAAC TTAAAAATnG 780  
 CTTTAAAATA ACAATTGChG GTGGTCAAGG CCATCTTAAA GGTCAAATTT TnAGAATTGG 840  
 TCATATGGGG AAAATTAGTC CTTTCGATAT TTTATCGGTA GTATCTGCTT TAGAAATTAT 900  
 45 TTAACTGAA CACCGTAAAG TTAATATAT CGGTAAAGGT ATATCAAAT ATATGGAGGT 960  
 TATTCATGAA GCAATTTAAT GTACTCGTTG CAGATCCCAT ATCAAAGAT GGTATCAAAG 1020  
 CATTATTAGA TCACGAACAA TTCAATGTAG ATATTCAAAC TGGCTTGTCC GAAGAAGCAT 1080  
 50 TAATCAAAT TATACCTTCA TACCATGCTT TAATCGTTCG TAGTCAAAC ACGGTTACTG 1140  
 AAAATATCAT AAATGCTGCT GATTCTTTAA AAGTAATCGC ACGCGCCGGT GTTGGTGTAG 1200

5  
 10  
 15  
 20

GTAATACGAT	TTCAGCTACT	GAACATACAC	TGGCAATGTT	ATTATCAATG	GCACGAAATA	1320
TTCCGCAAGC	ACACCAATCA	CTTACAAATA	AAGAATGGAA	TCGAAATGCA	TTTAAAGGTA	1380
CTGAGCTTTA	TCATAAAACA	TTAGGTGTCA	TTGGTGCTGG	TAGAATTGCT	TTAGGTGTTG	1440
CTAAACGTGC	GCAAAGTTTC	GGAATGAAAA	TACTAGCTTT	TGACCCTTAC	TTAACGGATG	1500
AAAAAGCAAA	ATCTTTAAGC	ATTACGAAGG	CAACAGTTGA	TGAGATTGCC	CAACATTCTG	1560
ATTTGCTTAC	ATTACATACA	CCACTAACAC	CTAAAACAAA	AGGCTTAATT	AATGCTGTCT	1620
TTTTTGCCAA	AGCAAAACCT	AGTTTGCAAA	TAATCAATGT	GGCACGTGGT	GGTATTATTG	1680
ATGAAAAGGC	GCTAATAAAA	GCATTAGACG	AAGGACAAAT	TAGTCGGGCA	GCTATCGATG	1740
TGTTTGAACA	TGAACCTGCA	ACTGACTCGC	CTCTTGTTGC	ACATGATAAA	ATTATTGTTA	1800
CACCTCATTT	GGGTGCTTCA	ACAGTCGAAG	CTCAAGAAAA	AGTGGCAATT	TCTGTTTCAA	1860
ATGAAATCAT	CGAAATTTTA	ATTGATGGTA	CTGTAACGCA	TGCAGTGAAT	GCACCTAAAA	1920
TGGACTTAAG	CAATATAGAT	GATACTGTAA	AATCATTTCAT	CAATTTAAGC	CAA	1973

## (2) INFORMATION FOR SEQ ID NO: 36:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 7620 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

35  
 40  
 45  
 50  
 55

GGTGTTCAG	ATGTCACTGG	TTGATTTT	ATTGTAGACG	GGTATTTTGG	GCTTCGCCA	60
TATTTATTTG	CCGGCTTACT	GTCAAAGCAT	AGGAATACTA	TCATAACAAT	TGTTAGGCCT	120
AAATGAACAA	AATAAAGAAG	TACTAACAAA	ATATTAAGAC	CCATCGGCAT	TAATGTAAAA	180
TCACTGTCAT	AATAACTATC	GATAATCTGT	AATACTATAT	AAAATATAAT	ACTGAATACT	240
GTCATAATCA	TTGGAAATAA	CATTGTTCTT	GATATATCGT	GAAATCTTCG	AACGCACAAC	300
GCTAAATTTG	GAATAAACGT	TGCCAAACTA	TAGACAAAAG	TATACACAGA	TGTAAGGATA	360
ATCATCAATA	TACTCATAAC	TATTAATGTT	TCGTTATCCG	CCGCTATAGA	AATAAAGAAT	420
AGAAATAGGT	TTATTATTAG	CACACACACA	GCTGGAACCA	TAAGTATCAA	ATGCCATAGT	480
GCCATATACC	AATATTCACT	ACGTCTTGAT	CTCCCCTTAA	AATTTACATA	ATTTTTCCAA	540
AATAAACGTA	ATGATTTTCAT	AAAACCTACT	TGAGGTAATT	GTTCCATTGT	AATCTCCCTT	600
TCGTTAATCA	TATTTATATT	TTTAATTATT	GTTACCGTTA	TAATTTACAA	GATTCATTAT	660

	GTAAATGAA AACCCGCTAC AAGTACACAT CTATATGGAG ACTCATTTGA AAGTCAACGC	780
	TTCGTTAACT ATACTAAAAA TATGTCATAC TGCAATGTTT ACGTTTAAAA GAGTCTCAAT	840
5	CTATGCAAAT AAAATATTCC ATAACAAAGT ATATACTTTA CATTTTATA ATTCTTAACA	900
	ATACTATTTT ATCAAACATT TACCACAATA AAAATATCTT TTTCAATTTT ATTTAAATTA	960
10	ATCATATAAT TGCGAGGAGA ATATTATGGA TTTGTTAAT AATGATACAA GACAAATTGC	1020
	TAAAACTTA TTAGGTGTCA AAGTGATTTA TCAGGATACC ACTCAAACGT ATACAGGCTA	1080
	CATCGTGAA ACGGAAGCTT ACTTAGGTTT GAATGATCGT GCGGCTCATG GCTATGGCGG	1140
15	TAAAAATAACA CCTAAAGTCA CGTCATTATA TAAACGTGGT GGTACAATTT ATGCACATGT	1200
	CATGCATACG CATTTACTCA TTAATTTTGT AACAAAATCT GAAGGTATAC CTGAAGGCGT	1260
	ACTTATCCGC GCAATTGAAC CAGAAGAAGG TTTATCCGCT ATGTTCCGTA ACAGAGGTAA	1320
20	GAAAGGCTAC GAGGTAACGA ATGGCCCAGG AAAATGGACT AAGGCATTTA ACATTCCACG	1380
	GGCTATCGAT GCGGCTACGT TAAATGACTG TAGATTGTCT ATTGATACTA AGAATCGTAA	1440
	ATATCCTAAA GATATTATTG CTAGTCCACG AATCGGTATT CCAAATAAAG GTGATTGGAC	1500
25	ACATAAATCT TTACGTTACA CAGTGAAAGG TAATCCATTT GTGTCTCGCA TCGGTAAATC	1560
	AGATTGTATG TTTCCCGAAG ATACTTGGA ATAAATGCCA TCTTTCATTG ATTACTATCA	1620
	TGAAATGAA ATCTATCTCC TTATAAGTCA ATCAATCGTG CCGTCAACAT GCGGATGGGT	1680
30	TGATTGTTTT TCTTTGTATC CATCATATTT TTTGATTCAT CTCCTCTTAT TGAAGTTGTT	1740
	CTTAATTATA AAATATAACA ATAGAATTAT TTATAATTAT TAAATTTAGA TGCATTAATA	1800
35	TTATTGATAT TATTTTCAA AACTAGAAAT ATTGATTGTT TGCATGTATA ATGTTAAAAG	1860
	CGCCCTTTTA TAACGCTTAC ATATAAAAGC TTATTTAGGG AGAGGGATAT TCAACAAGGG	1920
	GGATTGAAA ATGATAGAAC TTAATGCAAT TACAACATTA TGTTTAGCTT GTATCCTTTA	1980
40	TTTACTTGGT AAGGCTATCG TTAATCACGT TAATTTTTTA AAACGTATTT GTATACCAGC	2040
	ACCAGTGATT GCGGCTTAA TCTTTGCTAT TTTAGTTGCG GCTTTGGATT CATTTGGCAT	2100
	GGTTAAGATT AAATTAGATG CTTCAATCAT TCAAGATTTC TTCATGTTAG CATTCCTTAC	2160
45	GACAATCGGT CTTGGTGCAT CATTGAAATT ATTTAAATTA GGTGGCAAAG TCTTGCTATT	2220
	ATACTTTATG TTTTGTGCTA TCATTTCACT CAITCAAAC ATAGTTGGTG TATCACTAGC	2280
	AAAAGTATTA AATATTAAAC CTTTGTTAGG ATTAACAGCA GGTTCATGT CTATGGAAGG	2340
50	CGGTCATGGT AATGCTGCTG CTTATGGTAA GACAATTCAA GATTTAGGTA TTGATTCGGC	2400
	ACTGACAGCG GCTCTTGCG CTGCAACTTT AGGTCTTGTA TTTGGAGGGC TTATCGGTGG	2460
55		



	ATTTAAAGAT TATAGCCAAG TAGCATATAA CGAACATTTA CATAGTAAAT TTAATGCCAC	2580
	TGAAGTATTC TTCATTCAAT TTACAATCGT TGTATTCTGT ATGGCAGTTG GAAGTTATTT	2640
5	CAGTCATTTG TTTACAGCTC AAACAGGGAT TAATGTTCCA ATTTACGTTG GCTCATTATT	2700
	TGTAGCTGTT ATTGTCCGAA ATATCTCTGA AAGTTTAAAT TTAAATATTG TAGATTTAAA	2760
10	AATTACTAAT CAAATTGGCG ATGTCGCATT AGGTATTTTC TTATCTCTTG CGCTAATGAG	2820
	CATTCAATTA ATCGAAATTT ATAACTTGC TATACCTCTT ATTATTATCG TTTTAGTTCA	2880
	AGTTGTCGTT ATGATTTTAT TTGCTGTTTT AATTTTATTT AGAGGTTTAG GAAAAGATTA	2940
15	TGATGCTGCA GTAATGGTAG GTGGTTTTAT CGGTCATGGG CTGGTGCAc GCCAAATGCC	3000
	ATGGCAAAT TAGATGTTAT TACTAAAAAA TATGGAACT CACCTAAAGC ATATTTAGTT	3060
	GTACCTATTG TTGGTGCACT CTTAATCGAT TTAATTGGTG TTATAGTCAT TATGGGATTC	3120
20	ATACAATGGT TTAGTTAAAC ACCAACTCA TAAATAAAG AGGAGGCCTT CGCCTCCTcT	3180
	TTTATTTATC CTCGATGTAT ATTCAAGTTA CGTGTCTTA TCCATGACAA TATTTCCGGA	3240
	CTAAATACGA TTTGTTTTTG TGTTAAGTCG TCAATATTTT TAGCATCTAA CATCGTCATT	3300
25	ATTGATTTCA TGTGTCAAT AAATGATTCT ACATAAGCTA CTGTATGTGC AATGCCATTA	3360
	TTTTCAACTT GATTTAAAAA CGGACGTGAC ATACCAGTTG CCTTTGCACC AAGTGCTAAA	3420
	CTTTTAATTG CATCGAGTGG TGTACGTAAA CCACCACTCG CGAAAACTGA AATTTCGCTT	3480
30	TGATAAGCCG TTGTTTCAAG TAATGACTCA ACTGTAGACT GTCCCCATGA TGATAAGTAA	3540
	TCCATATCTT TATTTGCACG ACGTTCATTT TCAATATCTA CAAAGTTAGT ACCACCTTTG	3600
35	CCACTAACAT CGACATACTT GACGCCTATT TGTTGTAAGT CATGCATTAA TTCTTTGCTC	3660
	ATACCAAATC CAACTTCTTT TATAATGACT GGAACAGACA CTCGTGATAC AATCGACGCT	3720
	ATATTTATCTA ACCAAGTCAC AAATTCACGA TTCCCTTCAG GCATAACTAA TTCTTGAGGA	3780
40	GAATTAACAT GGATTTGTAA CGCTTGCGCC TCAAGTAATT CAACTGCTTC CAAAGCCTTT	3840
	TCTACTGGTA CGTCCGCACC AACATTGCTA AAAATCATGC CTTCAGGATT CATTTTTCGC	3900
	GCAATCGTAA ACGTCTCAGC CATGCGTGGA TTTCTCAATG CCGCATGTGT TGATCCAACT	3960
45	GCCATCGCTA AGCCAGTTTC TCTTGCAACT ACAGCTAGCT TTTCAATTGAT GTTTTTCGTC	4020
	CACTCGCTAC CACCCGTCAT TGCATTAATA TAAACCGGAT ATGCCATCGT TAAGTCAGGC	4080
	GTCTGTGATG TCAAATCGAT ATCATTTACA TTAATTGATG GGATAGAATG ATGCACAAAA	4140
50	CGCATCTTAT CAAAATCTGA ATGCATTGCG TCAGATTGGG CCATTGCTAT TTCAACATGT	4200
	TCATTTTTTC TCTGTTCTCT TTGAAAATCA CTCATGATTA AACCTACCTT TTCGTCATTT	4260
55		

	ATTACAGCTA AGCAAATATA ATATCCATAA TGAAATGTA ATGCCGGCAT ATTTACAAAG	4380
	TTCATACCAT AAATCCCAGC TATGAATGTT AACGGTGAAA ATATAACTGA TACTAATGTC	4440
5	AGTACTTGCA TAATACTATT CATTCTAAAT GACGTGTATG ACTCAAAATT TTCTCGTATT	4500
	TCGTTTGTCA TTTCTTGAGC AGTACGAATG ATATTACGTT GCTTAATCAA GTGGTCATCG	4560
10	ATATGTTGAA TGTATAGCGA ATGTTTATTA TCTATAATCA AATCACCATT TTGTTTCATT	4620
	GTATCAATTA GCTCTTGCAAT AGGAAACAGT ACACGTTTTA CTTTAATCAA ATCCGAACGT	4680
	AACTTAAAGA CACTATCCAT GACCATTTTA TTAAAGCGAT CATCTACATG GCGGTCTTCA	4740
15	AAATGATAAA CACTATCTTC AAGTGCATAT ACAAAGTTGA AATATTTATC AACCATCATA	4800
	TCTAAAATTA ATATGACGAC ATCTGCACAA TCTAATTCTG CATCTAATGT ATTCATATAC	4860
	TTATAGACTA CTTTATTTAA TGATTCCAAC GTTTGATGAT GATATGTTAC TAATACATTG	4920
20	TCTTGTATAA AAATATTTAG TGCTATTGGT GAATAGTTTG ACCCCATAAT ACTATGGAAT	4980
	ACTAAGTATT GATAATCTTT ATAAGATTTA TATTTAGCTC GTGGCATACC GTTAATTGCA	5040
	TCATCCACTT CTAAATCATT AAAATTAAAA TGTGCTTTAA ACCATTCAAT TTCTTGTTC	5100
25	TTCGGTTTCA CAAAATCATA CCAAACAATA GTCGCATCTT TTGGTATCTC TTTGATATCA	5160
	TCAACTACTT TAAACGGTTC ATATGTAGTT TGATACCGTA TCTTTAAAGC CATCGATACT	5220
30	CCCCCTAAAT AACGAATTCT CTATTATTTT ATCATGAATT AAATAACGTG TATGTCTTAA	5280
	TTTATTTTAG TATGATAGTC ACTAAGGAGA TGTTATTAT CAAACAACCT TTTACACATA	5340
	CTCAAACCGT AACATCTGAA TTCATTGACC ATAACAATCA TATGCATGAT GCAAATTATA	5400
35	ATATCATTTT TAGTGACGTC GTGAATCGTT TTAATTACAG CCACGGTCTT TCTTTAAAAG	5460
	AACGCGAAAA TTTAGCATAT ACGCTATTTA CACTAGAAGA ACATACGACA TACCTCTCAG	5520
	AATTGTCTCT TGGCGATGTA TTTACTGTTA CTTTATATAT TTATGATTAC GATTATAAGC	5580
40	GGTTGCATTT ATTTTAAACA TTAATAAAG AAGATGGTAC ACTAGCATCA ACAAATGAAG	5640
	TAATGATGAT GGAATTAAT CAGCACACAC GTCGTTCTGA TGCTTTTCCT GAATCATTTT	5700
	CAACACAAAT AGCACACTAT TATAAAAATC AATCAACTAT CACTTGGCCT GAACAATTAG	5760
45	GACATAAAAT AGCAATTCCA CACAAAGGAG CATTAAAATG ACAGATGCAT TACAACAAAA	5820
	GATTCATATC GAATTACTAG ATTTATTAGA TGATGTTAAG TTTGAATTAA CAGAATTAAA	5880
	TGCACAAAAA GGGTTATACA TTAACGGACC AGCAAATCAG CTACTTAAGC GTGGCGTGCA	5940
50	TATGGCTTAT GTTCAAGGAC AAAAGCAAGC CATCGATAAT ATTATGACTA TTGTGGAACA	6000
	ACAGCTTGAA AGATCAACAT TTCCTAGAAC ATTATGATAA ATTTCAAAAT GAGGTTGCTC	6060
55		

5  
 10  
 15  
 20  
 25  
 30  
 35  
 40  
 45

ATAATTTTTT AGATCAATTT TATCAAATTA AAGGGCAATA CTTTATCATC ACACATATCA 6180  
 ATACACTTAT TGGTGATTTT CACTCAGAAG CTCATTAACA ATTAGTCTAT ATAACCCTTG 6240  
 CTATATTTTC AAAAACAAAA CCCAATTACG TTTTCATGTC AAATATCATC TTGCATGAAA 6300  
 TCGTAACTGG GTCATTTATA TGTTATTAGT TATTTTGTGT TACATCCTCA TCTATCGATT 6360  
 TGGCAATTTG TTTAATAGCT TTATGTGATT GTCTAATTGG ATAAATTGGA AAATCATGTA 6420  
 CCATCTTAGG ATAATCATAA AACTCAATGT ATTGATGATG TTGCAACATC ATTTGTTCAA 6480  
 ATAGCTTCAT ATCAGGATGT GTCATTTTAC GTCCACCACC AAACATATAA ACTGGTGGCA 6540  
 ATCCTTCTAT TGTGCCATTA ATTGGCGATA TGGCCTTATC TGTTAATGGT AGGCCATTGC 6600  
 CCCATTTTTT CATAATCTCA TTGACACCAA ACTGACTTAG aACCGCATCT TGTTTCGATTA 6660  
 AGGCGTCCGA AATATCTTTA TTAGATAGTG TTGCATCTAA AATTGGTGAG ATTAAATACA 6720  
 ATTTATTCGG TAATGGCTGT TGATTAKCTA AAAGAGATTG TACAAAGGAT AATGCCAGTG 6780  
 CACCACCTGA ACCATCACCC ATGACTACGA CATTTTGATG TCCTACTTCA GATACTAATT 6840  
 GaTCATAAAC AC GTTGTATC GCTTGgNAAA GTATCGTCaA TATGnAAACT CTGGTGTCTT 6900  
 TGGATAGATA GGCAGTACAA CCTCATATAA TGLACTTAAa GTGATTTTAT CCCAACAAATC 6960  
 TCCAATGGAA CGGTGATGGT TGTAGTGCAT TGAATCCACC GTGAATATAT AAAATTTTCT 7020  
 TATCAATTTG ATGTCTGAAA TTAAAGCGAA AGACTTGCAT ATCATCTAAT GACAATTTTT 7080  
 CTAAATTTGC TTTAACATTT AATGTTGAAG GCTGCTTATG TTTTTTTCTA TTTTCAATTT 7140  
 CTCTTTTATA AAAAATCTT TCAACATCTT GATCATTTTT AAACATAATC GAGCGATTGT 7200  
 GAAGCAAATA TTTATTGACA ACGCTATTCA TAACACGGTT TCTAATCAAT GTCTTAACCT 7260  
 ACCTTTATAT ATTTTATGTA TCCAATGATk GTCTATCCCC TACATTCTTT GCCAAAAAA 7320  
 GTATATAATG TAGAAGATAT TTTCTTTTTT ACTTTCAAAT TTAAGACTAC AATTGAACAG 7380  
 TGATTTTTTCA TCATTATAAC AGACAAC TAG ACATATTGAT AAGTAAAGAA AAGAACTTTA 7440  
 TACGGAGGTA CCTTGCATGA CAAATCCAAA TCAACGATTA GAACCATTG ATGAGACATT 7500  
 TCAACAACCG AATATTCATC GTGGTAAGCG ATATGGTAAG AAAAAACGTT CATTGGTAAG 7560  
 CATGATTATT CAAATCATTG TTGTWATATT AACCACCATC GCTGGAATAC AGCATGGTGG 7620

(2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9834 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

	GTCATtACCG amTTTctTAG AaTCATTTAA AGATGATAAA TATACAAACG TTGGTAATTT	60
5	AAAAGAAGTG AATTTTGATA AAATTGCTGC GACGAAACCC GAAGTAATCT TTATCTCTGG	120
	ACGTACAGCT AATCAAAAGA ATTTAGATGA ATTCAAAAAA GCTGCACCTA AAGCGAAAAT	180
	TGTTTATGTT GGTGCAGATG AAAAGAACTT AATTGGTTCA ATGAAACAAA AACTGAAAA	240
10	TATCGGAAAA ATTTACGATA AAGAAGATAA AGCTAAAGAA TTAAATAAAG ATTTAGATAA	300
	CAAAATTGCT TCAATGAAAG ATAAAACGAA AAACCTCAAT AAAACTGTTA TGTATTTACT	360
	AGTTAACGAA GGTGAATTAT CAACATTTGG ACCTAAAGGT CGTTTGGTG GATTAGTTTA	420
15	CGATACATTA GGATTCAATG CAGTTGATAA AAAAGTAAGT AATAGCAATC ATGGACAAAA	480
	TGTTTCTAAC GAATATGTTA ATAAAGAAAA TCCAGATGTT ATTTTAGCGA TGGATAGAGG	540
20	TCAAGCGATA AGTGGTAAAT CAACTGCGAA ACAAGCATT AATAATCCTG TATTAAAAAA	600
	TGTTAAAGCA ATTAAAGAAG ACAAAGTATA TAATTTAGAT CCTAAATTAT GGTACTTTGC	660
	AGCTGGATCA ACTACAATA CAATTAAACA AATTGAGGAA CTTGATAAAG TTGTAAATA	720
25	ATTTTAAAG AGGGGAACAA TGGTTAAAGG TCTTAATCAT TGCTCCCCTC TTTCTTTAA	780
	AAAAGGAAAT CTGGGACGTC AATCAATGTC CTAGACTCTA AAATGTTCTG TTGTCAGTCG	840
	TTGGTTGAAT GAACATGTAC TTGTAACAAG TTCATTTCAA TACTAGTGGG CTCCAAACAT	900
30	AGAGAAATTT GATTTTCAAT TTCTACTGAC AATGCAAGTT GGCGGGGCCC AAACATAGAG	960
	AATTTCAAAA AGGAATTCTA CAGAAGTGGT GCTTTATCAT GTCTGACCCA CTCCCTATAA	1020
	TGTTTTGACT ATGTTGTTTA AATTTCAAAA TAAATATGAT AGTGATATTT ACAGCGATTG	1080
35	TTAAACCGAG ATTGGCAATT TGGACAACGC TCTACCATCA TATATTCATT GATTGTTAAT	1140
	TCGTGTTTGC ATACACCGCA TAAGATTGCT TTTTCGTTAA ATGAAGGCTC AGACCAACGC	1200
40	TTAATGGCGT GCTTTTCAAA CTCATTATGG CACTTATAGC ATGGATAGTA TTTATTACAA	1260
	CATTTAAATT TAATAGCAAT AATATCTTCT TCGGTAAAAT AATGGCGACA scgTGTTTCA	1320
	GTATCGATTA ATGAACCATA AACTTTAGGC ATAGACAAAG CTCCTTAACT TACGATTCCT	1380
45	TTGGATGTTT ACCAATAATG CGAACTTCAC GATTTAATTC AATGCCAAAT TTTCTTTGA	1440
	CGGTCTTTTG TACATAATGA ATAAGGTTTT CATAATCTGT AGCAGTTCCA TTGTCTACAT	1500
	TTACCATAAA ACCAGCGTGT TTGGTTGAAA CTTCAACGCC GCCAATACGG TGACCTTGCA	1560
50	AATTAGAATC TTGTATCAAT TTACCTGCAA AATGACCAGG CGGTCTTTGG AATACACTAC	1620
	CACATGAAGG ATACTCTAAA GGTGTTTAG ATTCTCTACG TTCTGTAAAA TCATCCATTT	1680
55		

## EP 0 786 519 A2

	AGTGTTCCTTT TTGAATAATG CTATTACGAT AATCTAACTC TAATTCCTTT GTTGTAAGTT	1800
	TAATTAACGA GCCTTGTTG TTTACGCAAA GCGCATAGTC TATACAATCT TTAACCTCGC	1860
5	CACCATAAGC GCCAGCATTC ATATACACTG CACCACCAAT TGAACCTGGA ATACCACATG	1920
	CAAAATCAAG GCCAGTAAGT GCGTAATCAC GAGCAACACG TGAGACATCA ATAATTGCAG	1980
	CGCCGCTACC GGCTATTATC GCATCATCAG ATACTTCGAT ATGATCTAGT GATAATAAAC	2040
10	TAATTACAAT ACCGCGAATA CCACCTTCAC GGATAATAAT ATTTGAGCCA TTTCCTAAAT	2100
	ATGTAACAGG AATCTCATT TGA TAGGCAT ATTTAACAAC TGCTTG TACT TCTTCATTTT	2160
	TAGTAGGGGT AATGTAAAAG TCGGCATTAC CACCTGTTTT AGTATAAGTG TATCGTTTTA	2220
15	AAGGTCATC AACTTTAATT TTTTCATTTG GGATAAGTTG TTGTAAAGCT TGATAGATGT	2280
	CTTTATTTAT CACTTCTCAG TACATCCTTT CTCATGCTCT TAATATCATA TAGTATTATA	2340
20	CCAATTTTAA AATTCATTTG CGAAAATTGA AAAGAAAGTA TTAGAATTAG TATAATTATA	2400
	AAATACGGCA TTATTGTCGT TATAAGTATT TTTTACATAG TTTTTCAAAG TATTGTTGCT	2460
	TTTGCACTC ATATTGTCTA ATTGTTAAGC TATGTTGCAA TATTTGGTGT TTTTTGTAT	2520
25	TGAATTGCAA AGCAATATCA TCATTAGTTG ATAAGAGGTA ATCAAGTGCA AGATAAGATT	2580
	CAAATGTTTG GGTATTCATT TGAATGATAT GTAGACGCAC CTGTTGTTTT AGTTCATGAA	2640
	AATTGTAAA CTTCGCCATC ATAACCTTCT TAGTATATTT ATGATGCAA CGATAAAACC	2700
30	CTACATAATT TAAGCGTTTT TCATCTAAGG ATGTAATATC ATGCAAATTT TCTACACCTA	2760
	CTAAAATATC TAAAATTGGC TCTGTTGAAT ATTTAAAATG aTGctACCGC CAATATGTTT	2820
	TGTATATTTT ACTGGGCTGT CTAAGAGGTT GAATAATAAT GATTCAATTT CAGTGTATTG	2880
35	TGATTGAAAA CAATTAGTTA AATCACTATT AATGAATGGT TGAACATTTG AATACATGAT	2940
	AAAGTcCTTT GATATTGAAA ATTAATTTAA TCACGATAAA GTCTGGAATA CTATAACATA	3000
40	ATTCATTTTC ATAATAAACA TGTTTTTGTA TAATGAATCT GTTAAGGAGT GCAATCATGA	3060
	AAAAAATTGT TATTATCGCT GTTTTAGCGA TTTTATTTGT AGTAATAAGT GCTTGTGGTA	3120
	ATAAAGAAAA AGAGGCACAA CATCAATTTA CTAAGCAATT TAAAGATGTT GAGCAAAAAC	3180
45	AAAAAGAATT ACAACATGTC ATGGATAATA TACATTTGAA AGAAATTGAT CATCTAAGTA	3240
	AAACTGATAC AACTGATAAA AATAGTAAAG AATTTAAGGC ACTACAAGAA GATGTTAAAA	3300
	ACCATCTCAT ACCTAAATTT GAAGCATATT ATAAGTCAGC AAAAAATTTG CCTGATGATA	3360
50	CAATGAAAGT TAAGAAATTA AAAAAAGAAT ATATGACGCT TGCAAATGAG AAGAAGGATG	3420
	CGATATATCA ATTAAAAAAA TTCATAGGTT TATGTAATCA ATCTATCAAG TATAACGAAG	3480
55		

	AATTAGCTGA TAATAAAAGT GAAGCAACTA ATCTTACGAC AAAATTAGAA CATAATAATA	3600
	AAGCGTTAAG AGATACTGCG AAGAAGAACC TAGATGATAG TAAAGAAAAT GAAGTAAAAG	3660
5	GCGCGATTAA AAATCACATT ATGCCAATGA TTGAAAAGCA AATTACCGAT ATTAACCAAA	3720
	CTAATATTAG TGATAAGCAT GTTAATAATG CAAGGAAAAA CGCAATAGAA ATGTATTACA	3780
	GTCTGCAGAA CTATTATAAT ACACGTATTG AAACAATAAA GGTTAGTGAG AAGTTATCAm	3840
10	AAGTCGATGT AGATAAGTTG CCGAAAAAGG GTATAGATAT AACTCACGGC GATAAAGCCT	3900
	TTGAAAAAAA GCTTGAAAAA TTAGAAGAAA AATAACTATA ATCATTTTTC AAAGTTAAAA	3960
15	ATTTTGAATT TATGGTTAAC ATGTCAACTT ACTATGTGTA TAATGGTAAA CATTGATATT	4020
	AACTATATGT ATAAAAATGT CACGCAGATG CTATTTAAAT GTGATAAATA TTTTtagagg	4080
	TGAATAGAGT GGCTATAAAG CTAAGTTCAA TTGACCAATT TGAACAGGTT ATTGAGGAAA	4140
20	ATAAATATGT TTTTGTATTA AAACATAGTG AAAGTTGTCC AATATCGGCA AATGCGTACG	4200
	ATCAATTTAA TAAATTTTTA TATGAACGCG ATATGGACGG TTATTATTG ATTGTCCAAC	4260
	AAGAACGCGA TTTGTCAGAT TATATTGCTA AAAAAACGAA CGTTAAACAT GAATCACCTC	4320
25	AAGCATTTTA TTTTGTAAAT GGTGAAATGG TTTGGAATCG AGACCACGGT GATATCAATG	4380
	TGTCGTCATT AGCACAAGCA GAAGAATAAT GAAACTATAG GGTTGGAACA TTTTGCCTTA	4440
	CACTACTAGA CGTGAATAGC ACAACTTAAA TTCGTGTGAA TCAGAGTAGT TTGGCTATAA	4500
30	TGATGTTCTG ACCTTTTATT TTATGTCACC TTTAGAAGCA GTTAAGTTAG TACTTTTTTA	4560
	CAAACATATG TATAATATAT TCGAGTATTT TTATTGAAAa tATTTTGGAA AACGACGAAT	4620
	CCAATAAGAA AATTAAACA TGATTTGTAA GTTAGTTTAA TAGGAAATAT ATGCTAAACC	4680
35	AAAAGAAGCA TATTGTTATT TACTGGAATA ATTAATAATC ATGTCATGTT AAATGTTAGC	4740
	ATATAATCAC GAGATAAAAT CTAAAATTTA AGATTAATCT TTTATGAATA AAAAACGTAT	4800
40	CACAACAAAT AATAAAGTAA GGTGGTCAAG GTTATGAAAG TATTAGTAGC CATGGATGAG	4860
	TTTCATGGAA TTATTTCAAG TTATCAAGCT AATAGATATG TTGAAGAGGC AGTTGCAAGC	4920
	CAAATTGAAA CTGCAGATGT AGTTCAAGTA CCATTGTTTA ATGGAAGACA TGAATTATTA	4980
45	GATTCTGTAT TTTTATGGcm ATCTGGGcaA AAGTATCGTA TACCAGTACA TGATGCAGAT	5040
	ATGAATGAAG TTGAAGGTGT TTACGGACAA ACTGATACAG GGATGACCGT TATCGAGGGG	5100
	AATTTATTTT TAAAAGGTAA AAAACCAATT GTTGAACGAA CAAGTTATGG TTTAGGAGAA	5160
50	ATGATTAAAC ATGCATTAGA TAACGACGCA AAACATGTTG TAATTTCACT AGGTGGGATT	5220
	GATAGTTTTG ATGCTGGTGC AGGTATGTTA CAAGCATTAG GTGCTCAATT CTATGATGAC	5280
55		

	GATATGTCGA ACTTACACCC TAAAATGGAA ACAGCAAGAA TTCAAGTAAT GTCGGATTTT	5400
	TCAAGTCGAT TATATGGTAA GCAAAGTGAA ATCATGCAAA CTTATGATGC GCATCAGTTG	5460
5	AATCATAATC AAGCAGCAGA AATCGATAAT TTAATTTGGT ATTTTAGTGA GTTATTTAAA	5520
	AGTGAATTGA AAATTGCAAT TGGTCCAGTT GAACGTGGTG GTGCTGGTGG TGGAAATGCA	5580
	GCAGTCTTGA ATGGACTGTA TCAAGCTGAA ATATTAACCA GTCATGCATT AGTAGACCAA	5640
10	CTAACACATT TAGAAAATTT AGTTGAACAA GCGGATTTAA TTATTTTTTG AGAAGGATTA	5700
	AATGAAAATG ATCAGTTGCT AGAAACGACA ACATTGCGTA TTGCAGAACT TTGTCATAAA	5760
	CATCAAAAGG TTGCCATTGC AATTGTGCA ACTGCTGAAA AGTTTGATTT ATTTGAATCA	5820
15	CAAGGGGTTA CAGCAATGTT TAATACATTT ATCGATATGC CAGAACTTA TACTGACTTT	5880
	AAAATGGGtT ACAAATTAGG CATTATACGG TTCAGTCTTT AAAACTGTTG AAAACACATT	5940
20	TTAATGTTGA GGTTTAGTAA AGAAGGACTA AATTGGTGAT GCTGTCATGA TGGTTAATAA	6000
	CATTTATGAT GGTTAGCAAA ACGAATTAGA AGATCGAAAG TATACGTAAA AAATATGAAA	6060
	AATCACGCTA TCATTGCACT GAATGTTAGC GTGATTTTTA TATATTAATT AAGCCTGAGT	6120
25	TGAACTAGTA TATAATCGTT GGTTTTAGT GATTTTCAGC GATATCTTCT ACAATTCCAA	6180
	TGATTACTTG TACTGCTTTT TCCaTAACAT CAATGGATGC aTATTCATAT GGGCCGTGGA	6240
	AGTTACCGCA ACCTGTAAAG ATGTTTGGAG TTGGTAACCC CATAAATGAC AATTGTGAAC	6300
30	CATCTGTACC ACCGCGAATA GGTTCAGTGT TTGCTGGAAT ATCTAATTTG GCAAAGACAC	6360
	GTTTAGGTAT ATCAATAATA TGAGGCAATG GTAATATTTT TTCTGCCATA TTGAAATATT	6420
	GATCCGATAT ATCAACTTTA ACTGGATAAT TTTCAAAATG GGCATTGATA TCGTCACGTA	6480
35	TTTCTAAAAT ACGTTTCTTA CGCAATTCGA ATTGTTTTTT ATCATGATCA CGAATAATGT	6540
	ATTGCAAAGT TGCTTTTTCA ACAGTTCCTT CAAAGTTCAT TAAGTGATAA AAGCCTTCGT	6600
	ATCCTTCTGT TCGCTCCGGA ACTTCACTAT CAGGTAGCAA ACTATCGAAT TGTTACCTA	6660
40	AACGTATTGC GTTTACCATT GCATTTTTAG CTGAACCAGG ATGAACATTT ACACCGTGGC	6720
	ATGTAATAAC CGCTTCAGCA GCGTTAAAGC TTTCATATTG TAATTCTCCA TATTGACTAC	6780
45	CATCCATAGT ATAAGCAAAA TCAGCATTGA AGCGGTCAAC ATCAAATTTA TGTGGACCAC	6840
	GACCGATTTT TCGTCTGGT GTAAATCCAA TCGGAATGGT ACCATGTTTA ATTTCTGGAT	6900
	GTTCTTGTA ATAACAAATA GCTTCATAA TTTCCACAAT ACCCGCTTTA TCGTCTGCAC	6960
50	CTAGTAACGA TGTACCATCA GTTACCATTA ATGTATGACC AACTAAACTG TTAAGTTCTG	7020
	GAAATACTTT AGGATCTAAG ACACGTTTAG TATTGCCTAG TTTGTATGGC TTACCATCAT	7080
55		

	GCGCCAAAAA	TCCAACTGTT	GGGACGTCGA	CATCGATGTT	ACTTTCTAAT	GTAGCAAATA	7200
	AGTAGCCATT	TTCATCTAAA	TCAGTTGGCA	ATCCTAATTG	TTGTAATTCT	TTTTCTAATA	7260
5	AATGTAACAA	ATCCCATTTG	TTTTCAGTTG	AAGGTGTTGT	TGTAGATTTT	GGATCAGATT	7320
	GCGTATCAAT	TGTCGTATAT	CTTGTTAATC	TATCTATCAA	TTGGTTCTTC	ATTATATTCT	7380
	ACCCCTTAAA	CTCTATTATT	CATGTTGTAA	GATTTTTTAT	ATGTCTTACC	TTTGATTTTA	7440
10	CCATACAGTT	GTTTGATACG	TGTGTATAGG	TAATATAGAA	TTTCAGAAAC	TAATATACCG	7500
	AAAGCAATCG	CACCTGAAAT	CAGTGTAATT	CTAAAAATGT	ATTTACAGCA	CTTGATAAAT	7560
15	CATTTGATAC	TAAAAAACGA	GTCGCTTGAT	AAGCTGCACC	ACCAGGTAAT	AATGGTATAA	7620
	TGCCTGGCAC	TATGAATATA	ATTACCGGTC	GTTTATATCT	GCGACTCATA	GTATGACTCA	7680
	TTAAGCCTAA	AATTAAGCTT	CCCAAAAATG	AAGCGCCAAC	TTTTCCAAAC	TCTAAATCTA	7740
20	CCGTTAATTG	GTAAATCGTC	CATGCAATGG	CACCCACAAA	TCCACATGCT	ACTAAGAGGC	7800
	GTTTGGGTGC	ATTGAAAATG	ATAGAGAAAA	GTACTGTTGA	TATAAAGCTG	ATTGTAAAAT	7860
	GAAATAAATA	AAATAGCATG	CTTTAACAGT	CCTTCCTTAA	ATGATTAATA	AAACGATTGC	7920
25	GACACCAGCA	CCGATTGCGA	ATGCTGTTAA	TGCAGCTTCA	ACACCGCGAG	ACATACCTGC	7980
	AAGTAATTCA	CCCCTAATA	AATCTCGAAT	GGCATTGGTA	ATTAATATAC	CAGGGACAAG	8040
	TGGCATGACA	CTGGCTATAG	TAATGATATC	TTGATTGGTT	GCAATGCCTA	ATTTAGTAAA	8100
30	TGTGGCTGCA	ATGGATATGA	CCACAGCGGC	TGCAACAAAC	TCTGAGAAAA	ATTTAATTTG	8160
	TATATAGCGT	TGCACAAAGC	TGAATGTTAA	AAATGCGGAT	CCGCCAGCAA	TGACTGCAAT	8220
	CCAACAATCT	GATGCGACAC	CACCAAACAT	AAATAGGAAG	AAGCCACATG	CAATGGCAGC	8280
35	TGCAAGAAAA	TTGTTAAAA	AAGAATATTG	TAATGATGCA	TGCTGTAAAT	GAATAAATTC	8340
	AGATTAGCT	TCATCAATTG	TGAGTTCTTT	ATTTGATATT	TTACGTGAAA	GACTATTCTG	8400
40	TAAAGCGATT	TTCTCTAAAT	CTGTTGTACG	CTCTTGATCA	CGAATTAATC	TTGTACTTGT	8460
	TCGATCGTTT	AATGAAAAAA	TAATTGCAGT	TGAACTGACA	AAACTATATG	TATTATGAAG	8520
	ACCATAACTA	TGTGCGATAC	GGTTCATTGT	ATCTTCAACT	CGATATGTTT	CAGCACCTGA	8580
45	TTCaAGTAAA	ATTCTACCTG	CAATTAATAC	AACATCAATC	ACTTTGTTTT	CATCTATAAT	8640
	TGTGATTGAA	TCTGGCATAT	CAATTCACCT	CCAATGATAT	GTGTTATTTA	TTGAACAAT	8700
	TGAGTTTAC	AACTTGTTGT	TACAACCTTC	AATAGTGAGA	CTTTGTGTTA	GTATGATGAA	8760
50	CTTGATGTT	TCAAATTTAA	ATAAGAAAAA	CTGTTAATCT	TTGCTATTAT	ACTATGATTT	8820
	AATAATAGCA	AAGGATTAAC	AGTTTTGTCT	TTGTTATAAA	TTGATAATAG	GGTTAAACAT	8880
55							



	TTTACGCTGT GATTTTGGAT CGTCATCTGT TAAATAACCA ACACCGATAG AACTGACAA	9000
	TTTAATAACT TCTTTGTTTG GTAAATGGAA TGATGATTTT TCAACACCCG AACGAATAIT	9060
5	TTCAGCTAAT TTAACACITT GATCAAGTGA ATAATTGTGA ATGACAACTG AGAACTCTTC	9120
	GCCACCATTT CTAAAAATTT TAAATTGATT CGGCACATAG TTTTAAAGTA ATTGAGACAT	9180
	TTGTTTTAAT ACAGCATCAC CTGATTTGTG TGAGTAGGTA TCATTGacAT CTTTAAATCC	9240
10	ATCGATATCG ATTAATAATA ATGCGATACT TTGATGTTCT TTTTCAGCTT TTCGTGAAAT	9300
	TTCAITTTAAA TGTCTATCAA ATTCTTTTAC ATTACCTAAG CCTGTTAAGT AATCATATTT	9360
	ATCTTCGTTT TCATAACGAT TTACGAGTGA GAAGAAATGC CAAATATCGA CAAATGTTAT	9420
15	CGCTGAAGCT AAAGTGATAA TTAATGAAAT TGGTATTAAA ATGATAACTT CCGATAGTGT	9480
	GTAAATAGGA CTCACCTAACG CGACACCAAA TAAATGATT ATTGTAACAA CATTAGTAT	9540
20	TAATAATGAT AGCACATCAT TTTGTTTTAA AAATGGTCCA ATAGCACTTG TTAGTGCAGC	9600
	AATAACAATC AACGTAACAC CGTACATAAT CGAGTTGTGA AATACTACAA TTTCAACAAT	9660
	TGCTACAATT ACTGTGGCAG ATAATGTATA GACCATATTT GTAAATCTAC CTAAAAACAA	9720
25	TAAAGGAACG AATGTTAAGT GAATTAAATA ATCTTCACGA TAAGGGATAG GGTAGACAGA	9780
	TAATAATAAT GATACGATTG TCATTAAAAC AGTGACATAA GCCTTAGAAA AAAC	9834

(2) INFORMATION FOR SEQ ID NO: 38:

30	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 23439 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

40	TCTCAATCAG ATGAAAAATT GCATATCGTA GGTTTTACAG AAAGTGCAAA ATATAATGCG	60
	TCATCAGTCA TTTTCACGAA TGACGCTACC ATTGCCAAGA TCAATCCTAG ATTGACTGGA	120
	GATAAAATTA ATGCAGTTGT TGTACGTGAT ACAAATTGGA AAGACAAAAA ATTAAACCAA	180
45	GAGCTTGAAG CGGTAAGTAT TAATGACTTT ATTGAAAATT TACCAGGTTA TAAACCACAG	240
	AACTTAACAT TAACTTTTAT GATTTCAATC TTATTTGTCA TTTTCAGCTAC AGTTATAGGC	300
	ATTTTCCTAT ATGTCATGAC ATTACAAAAG ACGAGTTTAT TTGGCATATT AAAAGCTCAA	360
50	GGATTTACGA ATGGCTATTT GGCGAATGTG GTAATTTTCG AGACGGTCAT ATTAGCACTA	420
	TTTGGTACGG CATTGCGCTT ACTGTTAACA GCGGTTACAG GTGCATTTTT ACCTGATGCA	480

55

	TCTGTATTAG GAAGTTTATT CTCCATTTTA ACAATTAGAA AAATAGATCC GTTAAAGGCG	600
	ATTGGGTAGG AGGTGTAGCA AATGTTGAAA TTTGAAAATG TAACAAAGTC ATTTAAAGAT	660
5	GGGAATCGTA ACATTGAAGC GGTAAAGAT ACAAATTTTG AGATAAATAA AGGTGATATT	720
	ATAGCATTGG TTGGACCTTC TGGCTCTGGT AAAAGTACAT TTCTAACTAT GGCAGGTGCT	780
	TTACAAACAC CGACATCTGG GCACATTTTA ATCAATAACC AAGATATTAC GACAATGAAG	840
10	CAAAAAGCAT TGGCAAAAGT TAGAATGTCT GAAATAGGTT TTATTTTACA AGCTACAAAC	900
	CTTGTACCAT TTTTAACGGT AAAGCAACAA TTTACATTAT TGAAAAAGAA AAATAAGAAT	960
15	GTTATGTCTA ATGAAGACTA TCAGCAACTT ATGTCACAAT TAGGTCTAAC TTCATTGCTT	1020
	AATAAGTTAC CTTCAGAAAT TTCAGGTGGT CAGAAACAAC GTGTGGCGAT AgCaAAGCGT	1080
	TATATACGAA TCCGTCGATT ATTTTAGCGG ATGAACCTAC CGCGGCGTTA GATACTGAAA	1140
20	ATGCGATTGA AGTCATTAAA ATTCTACGTG ATCAAGCCAA ACAAAGAAAG AAAGCATGTA	1200
	TTATTGTTAC ACATGATGAA CGACTTAAAG CATATTGTGA TCGTTCATAT CATATGAAAG	1260
	ATGGCGTCCT TAATCTTGAA AATGAAACAG TAGAATAGTT TTATTAAGCC GGTACATCAT	1320
25	GTGCCGGTAT TTTTATGTTT ATGTAATTATT TGAATAAACT TTCACATTCA ATTAATAATA	1380
	ATTATTATCG AAAATCAGAA ATATTCCGTG AAATATAATA TTTTTTGTAG TAAAATGGCC	1440
	TCTAAGTATT CAATATTTAA ATATGGGGAT TGAATATAAA ATTATCGTAA TGGGGGTCAA	1500
30	TGGTTATGGA TTTATTGATA GGTACTTTAT TTTTATTTT GGTCTTAGTG ATTTTACAT	1560
	TATTTACATA TAAAGCGCCT AATGGTATGC GTGCCATGGG AGCATTAGCT AATGCAGCAA	1620
	TCGCAACATT TTTAGTGGAA GCATTTAATA AATATGTTGG TGGCGAAGTA TTCGGTATTA	1680
35	AATTTTTAGA AGAGCTAGGA GACGCTGCGG GAGGTCTAGG TGGTGTCGCT GCCGCTGGAT	1740
	TAACAGCATT AGCTATCGGT GTGTCACCAG TATATGCATT AGTTATAGCA GCCGCGTGCG	1800
	GTGGTATGGA TTTATTACCA GGTTCCTTTG CGGGTTATAT GATTGGATAT GTGATGAAAT	1860
40	ATACAGAGAA ATATGTGCCG GATGGTGTG ACTTAATTGG ATCGATTGTC ATCTTAGCGC	1920
	CATTAGCTCG TCTTATTGCA GTATTATTAA CGCCAGTAGT GAATAGTACA TTGATTGAA	1980
45	TTGGTGATAT TATCCAAAGT AGTACGAATA CGAATCCAAT TATCATGGGT ATCATTTTAG	2040
	GTGGTATTAT TACGGTTGTC GGCACAGCGC CATTGAGTTC AATGGCATTG ACAGCATTAT	2100
	TAGGTTTAAAC GGGTGACCT ATGGCTATTG GTGCCATGGC AGCATTTAGT TCGGCATTTA	2160
50	TGAATGGGAC GCTATTCCAT CGCTTAAAT TAGGTGATCG TAAGTCTACG ATTGCAGTAA	2220
	GTATTGAACC TTTATCACAA GCAGATATTG TATCAGCCAA TCCAATTCCA ATCTATATTA	2280
55		

	ATGCGACAGG TACAGCTACA CCGATTGCAG GATTTTTAGT TATGTTTGGA TTAAATCATC	2400
	CGACGACAAT TGTGATTAT GGTGTAGTAA TGGCGATTGT AGGTGCGCTT GCAGGTTATC	2460
5	TTGGTTCAAT TGTATTTAAA AAATATCCAA TTGTTACTAA GCAAGACATG ATTAATCGAG	2520
	GTGCAGTAGA CGCATAGCAT CATCATATTG AATAGTAAAA ACAAATAAAA CATAGTAACG	2580
	TGATTCAAGT GATGTAACAG TCGATAATGA GTCACGTTTT TTTATAGAAA AATACAAGAC	2640
10	ATAAAAATGT CATAATTTAT TGTGACAAA TATCATACTG TATAAACATT TATCATTTTC	2700
	TCAAGTACCT TTTACACGAT GGAATGAACT TACTTTTTAC GAAATTATGC GTATTTTATA	2760
	AACAAATATC ATTGATATAA CGGTAAATGT AAGCGTTTAC AACAGAAATA ACAGCATGCT	2820
15	ACGATATTTT TGTAAATTCA CTGATTCAAG TATTTTAAGT CAATATGAGG AGGGATGTTA	2880
	TGAGCGATTG TGAGAAAGAA ATTTTAAAAA GAATTAAAGA TAATCCGTTT ATTTACACAAC	2940
	GTGAACTTGC TGAGGCAATT GGATTATCTA GACCCAGCGT AGCAAACATT ATTTCAAGAT	3000
20	TAATACAAAA GGAATATGTT ATGGGAAAGG CATATGTTTT AAATGAAGAT TATCCTATTG	3060
	TTTGTATTGG CGCAGCGAAT GTAGATCGTA AGTTTTATGT GCATAAAAAAT TTAGTTGCAG	3120
25	AAACATCAAA TCCTGTAACG TCAACACGCT CTATTGGTGG CGTAGCAAGA AATATTGCTG	3180
	AGAACTTAGG TAGGCTTGCC GAAACGGTCG CTTTTTTATC TGCTAGTGGA CAAGATAGTG	3240
	AATGGGAAAT GATTAAACGA TTGTCCACAC CATTTATGAA TTTGGATCAT GTTCAACAAT	3300
30	TTGAAAATGC GAGTACAGGT TCATATACAG CTTTAATTAG TAAAGAAGGC GACATGACAT	3360
	ATGGCTTAgC AGATATGGAA GTGTTTGAAT ACATTACGCC TGAATTTTTA ATTAAGCGTT	3420
	CACACTTATT GAAAAAGGCT AAGTGCATTA TTGTAGATTT GAATTTAGGC AAAGAGGCAT	3480
35	TAAACTTCTT ATGTGCCTAT ACCACGAAAC ATCAAATCAA ATTAGTTATC ACCACGGTTT	3540
	CTTCCCCAAA AATGAAAAAT ATGCCTGATT CATTACATGC TATTGATTGG ATTATCACGA	3600
	ATAAAGATGA AACAGAAACA TACTTAAATT TAAAAATAGA ATCTACTGAT GATTTAAAAA	3660
40	TAGCTGCTAA ACGCTGGAAT GATTTAGGTG TAAAAATGT TATTGTGACA AATGGCGTGA	3720
	AAGAACTCAT TTATCGAAGT GGTGAGGAAG AAATCATTA GTCAGTTATG CCATCAAATA	3780
45	GTGTGAAAGA TGTTACAGGT GCAGGCGATT CATTCTGTGC TGCAGTAGTG TATAGCTGGT	3840
	TAAATGGGAT GTCTACTGAA GATATATTAA TTGCTGGTAT GGTAAACGCA AAGAAAACGA	3900
	TAGAAACGAA ATATACAGTT AGGCAAAACC TAGATCAACA GCAACTTTAT CACGATATGG	3960
50	AGGATTATAA AAATGGCAAA TTTACAAAAG TATATTGAGT ATTCTCGAGA AGTTCAGCAA	4020
	GCACGGGAGA ACAATCAACC GATTGTAGCA TTAGAATCAA CAATTATTTT GCATGGTATG	4080

55

	GCCATTCCAG CAACCATAGC CATTATAGAT GGCAAAATTA AAATTGGITT AGAAAGCGAA	4200
	GATTTAGAAA TACTGGCAAC TAGTAAAGAC GTTGCTAAAG TATCTAGAAG GGATTTAGCA	4260
5	GAAGTTATTG CGATGAAGTG TGTGGTGCT ACTACTGTAG CGACGACGAT GATATGTGCT	4320
	GCAATGGCTG GTATTCAATT TTTTGTTACA GGAGGTATTG GGGGCGTCCA TAAAGGTGCA	4380
	GAACATACGA TGGACATTTT AGCAGACTTA GAAGAAGTGT CTAAAACAAA TGTCCTGTGTT	4440
10	ATCTGTGCAG GTGCCAAATC AATTTTAGAC TTACCTAAGA CGATGGAGTA TTTAGAAACA	4500
	AAAGGCGTTC CAGTTATTGG ATATCAAACG AATGAATTGC CAGCATTCTT CACTCGCGAA	4560
15	AGCGGTGTTA AGTTAAACAAG TTCGGTTGAA ACGCCAGAAC GACTTGCTGA CATTCATTTA	4620
	ACAAAACAGC AGTTAAATCT TGAAGGTGGC ATTGTTGTG CTAAATCCAAT TCCATATGAG	4680
	CATGCCTTAT CAAAAGCATA TATTGAGGCA ATCATAAATG AAGCTGTTGT TGAAGCGGAA	4740
20	AATCAAGGTA TTAAAGGTAA GGACGCCACA CCGTCTTGT TAGGGAAAAT TGTAGAAAAA	4800
	ACGAATGGTA AAAGTTTAGC AGCAAATATA AAACCTGTTG AAAACAATGC GCGTTGGGT	4860
	GCTAAAATTG CTGTCGCTGT TAATAAATTA TTGTAGGTGA TGATACATGA ATATTTTATT	4920
25	CGCTATCACA GGGATAGCAT TTGCACTATT TGTTGCGTTT TTATTCAGTT TTGATCGTAA	4980
	AAAAATAGAC TTCAAAAAGA CGTTAATAAT GATATTTATT CAAGTGTTGA TCGTGTTATT	5040
	TATGATGAAC ACAACGATTG GTTTGACAAT TTTAACTGCA CTAGGTTTAT TTTTGAAGG	5100
30	GCTAATAAAT ATTAGTAAAG CAGGCATAAA TTTTGT TTTT GGAGATATAC AAAATAAAAA	5160
	TGGCTTTACG TTCTTTTTAA ACGTATTACT GCCATTAGTT TTTATTTCTG TATTAATAGG	5220
	CATCTTTAAT TATATTAAGG TATTACCATT TATTATCAAA TATGTAGGTA TCGCTATTAA	5280
35	TAAAATAACT AGAATGGGGC GCTTAGAAAG TTATTTTGCT ATTTCAACAG CAATGTTTGG	5340
	GCAACCAGAA GTATATTTAA CAATAAAAGA TATTATTCCA AGATTATCTA GAGCGAAATT	5400
	ATATACAATT GCGACGCTCTG GTATGAGTGC TGTTAGTATG GCAATGCTAG GTTCATATAT	5460
40	GCAGATGATT GAACCCAAGT TCGTAGTTAC AGCAGTAATG TTAAATATTT TTAGTGCCT	5520
	TATCATCGCC AGTGTAATCA ATCCCTATAA ATCTGATGAT ACTGATGTTG AAATTGATAA	5580
45	CTTAACGAAA TCCACAGAAA CTAAAACATT GAATGGAAAA ACAGGAAAAC CTAAGAAAGT	5640
	TGCCTTTTTC CAAATGATTG GTGATAGTGC GATGGATGGG TTTAAAATCG CTGTTGTAGT	5700
	AGCCGTAATG TTGTTAGCAT TTATTTTATT AATGGAAGCA ATTAATATCA TGTGTTGTTAG	5760
50	TGTTGGTTTG AACTTTAAAC AGCTTATTGG CTATGTGTTT GCACCAATCG CATTCTTAAT	5820
	GGGGATTCCA TGGAGCGAAC TGTTCCAGCT GGCTCTTTAA TGGCGACTAA ATTAATTACA	5880
55		

	CAAGGTATCA TTTCAGTTTA CTTAGTAAGC TTCGCTAATT TTGGTACGGT TGGTATCATC	6000
	GTAGGTTCAA TTAAAGGCAT TAGTGATAAA CAAGGAGAAA AAGTTGCATC CTTTGCAATG	6060
5	AGGTTGCTAC TTGGTTCAAC TCTAGCTTCA ATCATTTTCAG GATCAATCAT TGGCTTAGTA	6120
	TTGTAAATGA ATCGAAGTAC CTAAATTAAA TTCATGGCAA AGCTAAACCC CGTCACCAAG	6180
	TTGGCGCAAC AGCGcATgca TAACTTAGTG ACGGGGTTTT ATCATAACAA TCTACTTTTT	6240
10	CGTAGCCGTT TTTGAAATGT ATGTTGATGG TTTATCTTTT TCAAAAATTG TTAATCCCGT	6300
	TATATCTTTT TTATGTTTTG AAGGGACAAT GAAGCTAAGT ATATAAGCAA AGACAAAAGC	6360
	AACTGTAAAT GAAATGGTAG ATACATAGAA AGGTGAGTTA CCTTTGCCAA CACCATTATA	6420
15	GACATAAGCA AAGATGATAC CCAATATTAA TCCACAAATA ACACCGAATG TATTCGTACG	6480
	TTTAGTGAAA ATACCAACTG CAAATACACC AGCCAATGGA ACGCCGAATA ATCCAGTCAC	6540
20	AAACAAGAAT AAATCCCATA AGTCATTTGA ATTAGAAGCA ATTAAGTATA GTGACATTCC	6600
	AAAACCGAAA ATACCTGCAA TGATAATAAT GAAACGTGCA AAGTTAACTT CGTGTGCTC	6660
	GCTACCTTTT CCGAAGAAGC GTTGCTTAAT GTCGATTGAA ATACAAGCAG ATATAGAATT	6720
25	TAACTAGAT GAAATGGTAG ACTGTGCAGC GCGGAAAATG GCTGCAATAA GTAATCCTGC	6780
	TACAAATGGT GGCATCTCAG TCAAAATGAA ATATGGCACT ACAGATGATG TATTGAAGCC	6840
	TTTTGGTAAA ACAGCTTCAT GTGTATAAAA TGAATACAGC ATTGTACCCA TACCATAAAA	6900
30	TAAGGGTGCT GAAATTAAAG CTAGGATACC ATTTGTCCAT AACGATTTAT TTGTTTCTTT	6960
	TAACTATCA GAAGCTTGAT AACGCTGCAC GACGTCTTGA CTCGCTGTGT ATTGATACAA	7020
	GTTGTTGAAA ATATTTCCCTA GGAAAATAAT TGGAATGGCA GCTGCCGCAG TATTTAGTTT	7080
35	CCAATTGTCT GCACTAATTA ATTTTTTG TG CTCAATCGCA TCTGCAAAGA CAGTGCCGAA	7140
	ACCGCCTTTA ATGTTCAACA CACCTAGAAT AATAATAACT AAAGCGCCGC CTAATAAAAT	7200
40	GACGCCTTGA ATGAAATCAC TCCAAACCAC ACCTTCGAAA CCACCTAAAA ATGTATATAA	7260
	AATACATAGT AAACCAACGA GTGATGCAAC GATATAAGGG TTCATGTCTG ATACAGATGT	7320
	GATTGCTAAT GTTGGTAAGT AGATAACAAT TGCAACACGC CCTAAATGGT AAACGACAAA	7380
45	TAATAATGAG CCAATGACAC GTATGCTAGG GCCAAATCTA GCTTCTAAAT ATTCATATGC	7440
	AGATGTTACC TTTAACTTTT TAAAGAAAGG GACATAGAAA TAAATAAGTA ATGGAATAAT	7500
	TGCGACGATA GCAATGTTAC CAGCGATATA TGACCAATCT GTTAAAAATG CTTTCTCTGG	7560
50	TGTCGACATA AATGTAATCG CACTTAACGT AGTAGCATAA ATTGAAAAGC CAACTACCCA	7620
	AGATGGCAAG CGACCACTTG CGGTAAAGAA ACTATTGGTA CTTTGGCTCG CGCGCTTGGT	7680

55

	1	TGTGCCAAAT CCAACTTCTT TCATGGGCAA CATCCCCTTT ACAATGTATT GATTCTTTGA	7800
	2	TGTCTATAAA TCGTATTTTG CAATGAGTTG ATCTAATGTT TGTGATGTG CTTGTTAAA	7860
5	3	AGGTTTGAAA GGTCTTTTCG GTAATCCTGC ATCAATGCCA CGATGACGTA ATATTTCTTT	7920
	4	CAATGTTGGA TAAATCCCCA TTGATAACAC TGTTTCGATA ATGTCGTTTG AATCATGTTG	7980
	5	CAGTTGGTAA GCTTCTTGAA TTTGACCTTG TCGTGCTAAG TCGAAGATTT TTCTTGACAG	8040
10	6	GCGACCATTA ACGTTATATG TAGAACCAAT TGCACCATCT ACGCCAGAAA TCGTAGCTTG	8100
	7	AACTAACATT TCATCAAAGC CAGATAAGAT TAATTTGTCT GGAATGCTT TTCTAATACG	8160
15	8	TTGAGTAGG AAGAAGTTTG GCGCTGTATA TTTAACACCA ACAATTTTTT CATGATTAAA	8220
	9	TAGCTCGCTG AATTGTTCAA TAGAAATATT CACACCTGTT AAATCTGGTA TTGCATAAAT	8280
	10	AATCATATTG TTCTGAGTTG CTTGATAAT ATCGAAATAG TAATCTCTAA TTTCTTCAAA	8340
20	11	AGTAAATGGA TAGTAGAATG GTGTACGGC AGAAAGTGCA TCATAACCGA GTTCTGTGGC	8400
	12	ATATTTTCCA AGTTCAATGG CTTCAATTAA ATCTAACGAA CCTACTTGAG CAATCAATTT	8460
	13	CACTTTATCC CCAACTGCCT CTTTGGCAAC CTTGAAAAC TGTCTCTTCT GCTCTGTATT	8520
25	14	TAATAAAAAG TTTTCGCCG AGCTACCATT TACATAAAGA CCGTCTAATT CTTCAGTTTC	8580
	15	AATGGCATTG TGAGCAATTT GTTTAAGTCC TTGTTCAATT ACTTGACCAT TTTTATCAAA	8640
	16	AGGAACGAGT AACGCTGCAT ATAAACCTTT TAAATCTTTG TTCATTATGA AGTCCCTCCA	8700
30	17	AAAATCATTT GATAATATAG TTTACAGCTA TAATTGTAAA CGCTATCATA AAATGTAAAC	8760
	18	ATATCTTTTT GAAAATGTGA GTCATATTTA TGTATAATTA ATGAAAATGT TTTTCAAAAT	8820
	19	CAATAGAAAT GGAGTGAGTA AGGTGTATTA CATCGCAATC GATATTGGAG GCACTCAAAT	8880
35	20	TAAATCGGCA GTTATTGATA AGCAATTGAA TATGTTTGAC TATCAACAAA TATCAACGCC	8940
	21	GGACAACAAA AGTGAGCTTA TTAGTGACAA AGTATATGAG ATTGTAACAG GATATATGAA	9000
40	22	GCAATATCAG TTGATCCAAC CTGTCATAGG TATTTTATCA GCAGGCGTTG TTGATGAACA	9060
	23	AAAAGGCGAA ATTGTATACG CAGGGCCAAC CATTCCGAAT TATAAAGGTA CTAATTTTAA	9120
	24	GCGATTATTA AAATCACTGT CTCCTTATGT CAAAGTAAAA AATGATGTAA ACGCTGCATT	9180
45	25	ACTAGGCGAA TTGAAATTAC ATCAATATCA AGCAGAACGG ATCTTTTGTA TGACGCTTGG	9240
	26	TACAGGCATT GGGGGTGCCT ACAAGAATAA TCAAGGTCAT ATTGATAATG GTGAGCTTCA	9300
	27	TAAGGCAAAT GAAGTTGGGT ATTTATTGTA TCGTCCAAC TAAAAACAA CGTTTGAGCA	9360
50	28	ACGTGCTGCA ACGAGTGCAT TGAAAAAGCG CATGATTGCC GGAGGATTTA CGAGAAGCAC	9420
	29	ACATGTGCCA GTATTGTTG AAGCAGCTGA AGAAGGTGAT GATATTGCAA AACAAATATT	9480
55	30		

	AGGGCTTATA TTAATTGGGG GCGGTATATC TGAACAAGGA GATAATCTCA TTAAATATAT	9600
	CGAGCCGAAA GTTGCACT ACTTACCAA AGACTATGTT TATGCACCAA TACAAACGAC	9660
5	TAAGAGTAAA AATGATGCAG CATTATATGG CTGTTTGCAA TGATAGTTGA AAGAAGGAGT	9720
	CATTCTAAAA TAGAATTTGA AACCGTTACG AGAGATGAGA GCTGTTGTTA GTTCCACACA	9780
10	TCACACTCTA TCTAGGACCA ATCTAAACTA TATCAACCAA CAGTGTGCCA CGGGCAAATT	9840
	AAATTGAAGA AGCTGAGATA TTAATAATTT AGAAAATGTA AAAAAATATT TGGTATTGAA	9900
	ATTAAAAAAG CACCTAGCAA CTCGTTGGGA CAATCACGAT GATTGTCTAC AGTTGCAGGT	9960
15	GGATTTGAAT ATACTACTAG TTATTTGTTG TCTAGGATAA TAGATTTAGT ATGTTGATAA	10020
	GTTTGACTCA GATTGCTATT TTCTAATAAA TGATAACTCA CGATATCGAT TAAAAAGAGT	10080
	GTCGCAATTT GTGTGTTGAT AAATTGATGG TCGGTATTAC GCGATTGATC CGTTGTTAAA	10140
20	AGTACTAAAT CTGCACAATC TGTAAGTTTA CTACCTTCAA AATTGTGAT GGCAACGACA	10200
	TATGCACCAT GAGATTTGGC GACTTCCGCT GCAGAAATTA ATTCCGAAGT ATTACCACTA	10260
	TTTGACATAG CAATAACAT ATCCGAATGA GATAGTAGGG ATGCCGATAT TTTCATTAAA	10320
25	TGTGAATCGG TAGTAACATT ACCTTTTAGC CCCATACGAA TCATACGATA ATAAAAATTCA	10380
	GTCGCTGATA AACCAGAGCT ACCTAGTCCA GCAAAGAGTA TATGTCGACT TGATTGAAGT	10440
	TTGTCGATAA AGGTTTGGAT AATGTCGTTA TCAATAAATT CACCAGTTTG TTGAATGATT	10500
30	TGTTGATGAT ATTTATGAAT TCTTTGAATA ATTGGGCTAT TTTCAATAAC TGTCTCTGTC	10560
	ATTTCTTGTT GAATATTAAT TTTTAAATCT TGGAAATCT CATAATCCAG CTTATGACTA	10620
	AAGCGTGTC TCGTTGCTGG TGATGTACCA ATCGCATGGG CTAAGGAGTT AATCGTTGAA	10680
35	AAGGCATCGC TATAACCATT TTGTCTTATA TAATTGACGA TCGGTTTATC AGTTTTTGTA	10740
	AATAAATGTT GATAACGTTG AACACGATTC TCAAATTTCA TTGTGTCACC CCTTCATCTT	10800
40	AATGATTACT ATTATATATG AAAAAATTTT TCAAGATAGT AAAAAGCATT GATAAAAAATT	10860
	ATCTTAATGA TATATTGTAA ATGACTTTAC GTGAAAAAAC GACTTATGGA GTGAGGAATA	10920
	ATGTTACCAC ATGGATTAAT AGTATCTTGT CAGGCACTAC CAGATGAACC ATTGCATTCA	10980
45	TCTTTTATTA TGTGAAAAAT GGCATTAGCT GCGTATGAAG GTGGTGCTGT TGGTATTGCG	11040
	GCAAATACTA AGGAAGACAT TTTAGCAATT AAAGAAACGG TAGATTTACC AGTTATTGGC	11100
	ATTGTGAAAC GTGACTATGA TCACTCAGAT GTTTTCATTA CTGCAACGTC AAAAGAAGTT	11160
50	GATGAACTGA TAGAAAGCCA ATGTGAAGTC ATTGCATTGG ATGCAACGTT ACAGCAACGT	11220
	CCGAAAGAAA CGTTAGACGA ATTAGTATCA TATATTAGAA CACATGCACC GAACGTTGAA	11280
55		

TATATTGGCA CGACGTTACA TGGCTATACT AGTTATACGC AAGGACAATT ACTTTATCAA 11400  
 AATGACTTCC AATTTTTTAAA AGATGTACTA CAAAGTGTTG ATGCAAAAGT TATTGCGGAA 11460  
 5 GGTAATGTCA TTACACCGGA TATGTATAAA CGTGTGATGG ACTTAGGCGT TCATTGTTCA 11520  
 GTCGTTGGTG GTGCGATAAC ACGACCAAAA GAAATTACGA AACGTTTGTG TCAAATTATG 11580  
 GAAGATTAAA TGATAACGAT AAAAAAACGA GATGACCATC ATTAATTAAA GGCACCTAAT 11640  
 10 TATCTTAGGT GGCTGAATGA ATGTAATGGG TTCATCTCGT TTTGTTTGTT TATGATAGTG 11700  
 ATTTTATTTT CAACTTTATC CAAAAATAAG TAAAGCGACG GGGATGGTGA TTAATAGCGA 11760  
 CAACGCCACG CGTAAAAACC AAATGATGAT GAGTTTCCAG ACAGGTATTT TAATTTCACT 11820  
 15 TGCTAGTATA CATGGCACTA ATGCTGAGAA AAAGATAATG GCTGATACGC TTAATACACC 11880  
 GACGACAAAT TTAGTACTCA TTGCAGCTTT AGTTACTAAC AAAGATGGTA GAAACATCTC 11940  
 20 TACAATAGAA ACGCTGACGC TTTTGCTAGT AAAGCCTGAT CAGCAATTGG GAAAATATAA 12000  
 ATAAATGGAT AGAAGATATA GCCAAGCCAA TCAATGAATG GTGTATAGTT CGCTACAATC 12060  
 AGTCCTAAAA AACCAATCGA TAATATAGAA GGTAATAATC CAACAGTCAT TTCTAAACCG 12120  
 25 TCTTTCAAAT TGTCCCAAAC GTTCTTCACG AGAGATGGTG TTAATGCATT TTGTTTCATC 12180  
 GCCTCTGCAT ATGCAGTTTT CAGTCTGCTT CCTTCAATAG CAACTTCTTG TTCTCCTTCT 12240  
 TGTCCGTTAT AATATTCTGT TGATTCATTG CTGATTGGCG GTAGCCATGC AGTAATTGCA 12300  
 30 GTCACGACAA ATGTGATGAC TAAAGTTATC CAAAAGTATA AATTCCAATG CGGCATTAAT 12360  
 CCTAAAGTTT TAGCAACGAT AATCATAAAA GTTGCTGAAA CTGTTGAAAA GCCAGTCGCA 12420  
 ATAATCGTGG CTTCTCGTTT GTTGATACATC CCTTGCTTAT AGACACGATT AGTAATCAAT 12480  
 35 AATCCTAAGG AATAACTGCC GACAAACGAA GCCACTGCAT CGACAGCGGA TTTTCCTGGT 12540  
 GTTTTAAAAA TAGGTCTCAT AATAGGCTCC ATATAAACAC CGACAAATTC TAATAAGCCA 12600  
 TAGCCCACTA ATAAAGAAAG CGCAATTGCA CCTACTGGAA TTAAGATACT TAATGGCATC 12660  
 40 ATTAATTTTT CAAACAAAAA CGGACCATAG TTAGCTTTAA ATAGTATTGA TGGACCGATT 12720  
 TTAAATACAT ACATTATACC GATCATTGCA CCTGCAACTT TAAATAATGT AATGACCAAG 12780  
 45 TTTGTGATTG AAGTCATAAA AGTACGTCTC ACTATTGGTA ACGCTGTACC AATTAAAATC 12840  
 ATAATCAGTG CAACATAGGG CATAAGTGGA CCTATGATTG AGCGAATGGC TAGATGAACA 12900  
 TGATCGACGA AAATAGTGTT GTTACCATTA ATCGTAAAAG GAATAAAGAA ACATAGTATG 12960  
 50 CCCACTAAAC TATAGACAAA AAAACGCCAT GCACTTGGTT GTTGTGCATT AGAATGATAT 13020  
 TGATTCATTA AAGCAACCCC TTTGTTTAAA TGAATACACA AAAGTGTATG ATGCATCTTC 13080

55



	ATAGTTTGAA TTATTTTCAT ACCAATACAA ATTAAC TAAT TATATATAGA TTGAACTAT	13200
	ATTACTTAAT AAAATATTTA TCTTAAATGT TGTGTGTGTTG ATTCAACACC ACAACTAAAA	13260
5	GTGTTTATAA ATTATTTTGA AATACACATA TTTGTAAATG ATTAGTATCG ATTTAATATC	13320
	GTATTATTAA ATTTTATTATA ATTTTGTAGT CTTAATCMAA AAATAATATA TGTCATGTTA	13380
	TATTGAAGGT GCAGTTGTTT TTCATTCTCA AGAGGGGGTC AAAAAAATAC TTTTGAGGTG	13440
10	ATTATATGTT AAGAGGACAA GAAGAAAGAA AGTATAGTAT TAGAAAGTAT TCAATAGGCG	13500
	TGGTGTCACT GTTAGCGGCT ACAATGTTTG TTGTGTCTATC ACATGAAGCA CAAGCCTCGG	13560
	AAAAACATC AACTAATGCA GCGGCACAAA AAGAAACACT AAATCAACCG GGAGAACAAG	13620
15	GGAATGCGAT AACGTCACAT CAAATGCAGT CAGGAAAGCA ATTAGACGAT ATGCATAAAG	13680
	AGAATGGTAA AAGTGAACA GTGACAGAAG GTAAAGATAC GCTTCAATCA TCGAAGCATC	13740
20	AATCAACACA AAATAGTAAA ACAATCAGAA CGCAAAATGA TAATCAAGTA AAGCAAGATT	13800
	CTGAACGACA AGGTTCTAAA CAGTCACACC AAAATAATGC GACTAATAAT ACTGAACGTC	13860
	AAAATGATCA GGTTCAAAAT ACCCATCATG CTGAACGTAA TGGATCACAA TCGACAACGT	13920
25	CACAATCGAA TGATGTTGAT AAATCACAAAC CATCCATTCC GGCACAAAAG GTAATACCCA	13980
	ATCATGATAA AGCAGCACCA ACTTCAACTA CACCCCGTC TAATGATAAA ACTGCACCTA	14040
	AATCAACAAA AGCACAAGAT GCAACCACGG ACAAACATCC AAATCAACAA GATACACATC	14100
30	AACCTGCGCA TCAAATCATA GATGCAAAGC AAGATGATAC TGTCGCCAA AGTGAACAGA	14160
	AACCACAAGT TGGCGATTTA AGTAAACATA TCGATGGTCA AAATTCCCCA GAGAAACCGA	14220
	CAGATAAAAA TACTGATAAT AAACAACATA TCAAAGATGC GCTTCAAGCG CCTAAAACAC	14280
35	GTTGACTAC AAATGCAGCA GCAGATGCTA AAAAGGTTG ACCACTTAAA GCGAATCAAG	14340
	TACAACCACT TAACAAATAT CCAGTTGTTT TTGTACATGG ATTTTITAGGA TTAGTAGGCG	14400
40	ATAATGCACC TGCTTTATAT CCAAATTATT GGGGTGAAA TAAATTTAAA GTTATCGAAG	14460
	AATTGAGAAA GCAAGGCTAT AATGTACATC AAGCAAGTGT AAGTGCATTT GGTAGTAACT	14520
	ATGATCGCGC TGTAGAACTT TATTATTACA TTAAAGGTGG TCGCGTAGAT TATGGCGCAG	14580
45	CACATGCAGC TAAATACGGA CATGAGCGCT ATGGTAAGAC TTATAAAGGA ATCATGCCTA	14640
	ATTGGGAACC TGGTAAAAG GTACATCTTG TAGGGCATAG TATGGGTGGT CAAACAATTC	14700
	GTTTAATGGA AGAGTTTTTA AGAAATGGTA ACAAAGAAGA AATTGCCTAT CATAAAGCGC	14760
50	ATGGTGGAGA AATATACCA TTATTTACTG GTGGTCATAA CAATATGGTT GCATCAATCA	14820
	CAACATTAGC AACACCACAT AATGGTTCAC AAGCAGCTGA TAAGTTTGA AATACAGAAG	14880

55

	ATTTAGGATT AACGCAATGG GGCTTTAAAC AATTACCAAA TGAGAGTTAC ATTGACTATA	15000
	TAAAACGCGT TAGTAAAAGC AAAATTTGGA CATCAGACGA CAATGCTGCC TATGATTTAA	15060
5	CGTTAGATGG CTCTGCAAAA TTGAACAACA TGACAAGTAT GAATCCTAAT ATTACGTATA	15120
	CGACTTATAC AGGTGTATCA TCTCATACTG GTCCATTAGG TTATGAAAAT CCTGATTTAG	15180
10	GTACATTTT CTTAATGGCT ACAACGAGTA GAATTATTGG TCATGATGCA AGAGAAGAAT	15240
	GGCGTAAAAA TGATGGTGTC GTACCAGTGA TTTCGTCATT ACATCCGTCC AATCAACCAT	15300
	TTGTTAATGT TACGAATGAT GAACCTGCCA CACGCAGAGG TATCTGGCAA GTTAAACCAA	15360
15	TCATACAAGG ATGGGATCAT GTCGATTTTA TCGGTGTGGA CTTCTGGAT TTCAAACGTA	15420
	AAGGTGCAGA ACTTGCCAAC TTCTATACAG GTATTATAAA TGACTTGTG CGTGTGGAAG	15480
	CGACTGAAAG TAAAGGAACA CAATTGAAAG CAAGTTAAAT TCATCTCTG AATTTAATAT	15540
20	GCTATGTAAA TCGTGCTGTT ATCATGGCAC ATCAGATATA AGTAGCATCA CAGTGTGAA	15600
	TTTAAAAATA GTAAAGTGAA ATAAAGCGCC TGTCTCATT GCGAAACTA AAGGGACAGG	15660
	CGTATCTGTT TATGAGCTTA ATAAATTGTA TGAATAATAT GGTGATCGA ATAAGTGT	15720
25	ATCATGATGA TAAATTGAGT TTTTAAAT AATGATATAT TACATCATTG TTATAGCGTT	15780
	TAAGAAATCA ACAACTTTAC GATAAATAGT GATTGCTTCG TCATTAGGTC TACGATCAAA	15840
	ATCATGCTCG TTTTATTCA CGCGTTCAA TGTGAATGT GGAACATGAT TCATGATATG	15900
30	TTGCTTTCC TCAACGGGAA CATCATAATC GCCATTACAA TCGCAATGA AAACAGGTGG	15960
	AAGTGTTTTA AGTTCATCTG GTGCAATATT ATATTTTGAA TTAGTATAAT CAGCAATGTT	16020
	AATCATATTT ATCCATTAC CTGTGCCACG TGCATAAAGC TAGATTAAAA AACGTTGTGC	16080
35	GATTTGATCT TGAACAACCG GTGTTGGTGA AGTGAGTTGT GCAATCATTG TTTCGTTTAC	16140
	GCTTTGAGCT ATTTTTCGCT AATAACTATT AGTTGTTTAA AAAGGTTGAG TGTGATGCG	16200
40	ACTATAACCA TAAAAATCAA TAACACCATC AATATCTCTG TCTCGTGCAA TTAATAGACT	16260
	TAAATATGCA CCTGATGATC TGCCAAAGGT AAAAATAGGG CAATTAGAAT ATTGTGATTG	16320
	AATCGCATCG AATGATGCgn AGNACATCCT CAATAATGCA ATCGAGACTT ACTTCTGGTA	16380
45	ATAAACGATA ACTTAGTTGA ATTAAATCGT AATGTTCCGT AAgATATCGA TATACTGTGG	16440
	GGATAAATCG TTAGCTTTAC CGAACATTAA TCCACCACCG TGGATGTAGA CAATAGCGCC	16500
	TTTGTGTTGG TGATTTTGT CTTAATAAT TGTGTAAGGT AATGCAAATG CATCTTTAGT	16560
50	AATTACTTTA TCTTTAATTT CAGTCACGAT TTAATAGGCT CCTTATTTT GATATTGATG	16620
	TCATTATAAC ACTGTCTTAA ATTTCCATGA AAAATAGTCT TAAGACGATG AGTCATGATA	16680
55		

	CATCAITTTA	ACAATATCTT	TAAAAGCAGC	ATGTGGAATG	GCTAAATCTT	CTAAATCTGC	16800
	CATAGAAAAT	TCAAGATTGA	TATCATGTGG	TCGCTGTTCA	GCAAGTTTAT	GCACAAAAGTC	16860
5	AGGTTCTGTG	ACAAAAGGCG	AAGACATGCC	GACCATATCT	GCATGTTGTA	AAGCATCTAA	16920
	AGCAGACTCT	GGAGAATTAA	TCCCGCCACT	TGCAATTAAA	GGGATACGAC	CTGCTAAATG	16980
10	TTCATAGACA	ATTTGGTTAA	CTGGTCGACC	GAAATGATCA	CCTGGTGTA	GAGACGTATT	17040
	TTGATAAATA	TGTCGACCCC	AGCTAGCGAT	TGCTAAGTAT	TGGATGTTTG	AAACGTCCAT	17100
	GACCCAATTG	ATTAATTGGT	TGAACTCGTC	AATGGTATAT	CCTAAATCAC	TGCCTCTGGT	17160
15	TTCTTCTGGC	GTTGCTCGAA	ATCCTAAAAT	AAAATTGTCA	GGTGCTTCTT	TATCAATCAC	17220
	TTCTTGTA	GCACGCATAA	CTTCTAAACA	TAATCTTGCA	CGATTTTITA	ATGAGTCGGC	17280
	ACCGTAATGG	TCTGTACGTT	TATTGCAAAA	AGTTGAGAAA	AATGTTTGAA	TCAGCAAACG	17340
20	TTGTGCAATC	GAAATTTCCA	CACCATCAAA	ACCTGCTTTA	ATCGCGCGTA	ATGTAGCATC	17400
	GCGATACTGC	TGAATGATGC	TATTGATTTT	CTCATGAGAC	ATGGCGATAA	CATCGTGTTT	17460
	AATCGGTGAA	TGCAATGTCA	TAGGGCTTGG	TCCATACACC	TTTCCAAAAT	TTAAAATGGC	17520
25	TTGATTTGAA	AAACGACCAG	CATGCGCTAG	CTGGATAATA	GCGAGGCTAC	CATGTTGTTT	17580
	CATCGTAGAT	GCCATGTTAG	TTAATCCAGG	GATACAAGCA	TCATGATCAA	TATTAAAGCC	17640
	ATATTCAAAC	AATTGACCAT	AAGGTTCAAT	GTAAGCAGCG	CCGGTGACTT	GCATTCCAGC	17700
30	TGAATTAGAG	CGACGTGCAG	CATAAGCCAA	GTCTTCTTTT	GTAATATAGC	CTTCTTTTGT	17760
	TGATGTGTTT	ACGGTCATTG	GTGATAATAC	AAAGCGATTG	GAAATTTTGA	TGCCATTAGG	17820
35	TAAGTGGATT	GATTGTAAAA	GTGGTTTGTA	TCGGTACATA	CTATGATTCC	TTTTCTATTG	17880
	AATATTGTTT	TCAAAGTACC	ATGAAAGAA	TGAATAATCA	ATGATGAACA	GTCTTGATAG	17940
	AATAGAATTG	GTACATGGAA	AGTATTTTAA	AAATTAACT	AATGAATGGC	ATTTGTAGGT	18000
40	CTGAAAATAT	GAATATGAAA	AAGAAAAATA	AAGGCGAAAA	GATATAAAAG	TTAATTGAAA	18060
	AACGTTATCA	TATACGTGGG	TATATGAAGA	GGGAATGGTA	TTAAGAACGC	TAAAATGTTA	18120
	TGTCGGTTTG	ACATGACAGG	ATAAGTTTGG	AGATGACGGA	TTGGTTAAAT	TAAGCGTATT	18180
45	AGACTATGCC	TTAATAGATG	AAGGTAAGGA	TGCACAAAAG	GCATTGCAAG	ATTCAGTGAC	18240
	ACTTGCAAAA	TTAGCAGATC	GACTTGGCTT	TAAGCGAATT	TGGTTTACGG	AACATCATAA	18300
	TGTACCAGCG	TTTGCGTGTA	GTAAGTCCAGA	ACTTTTGATG	ATGCATACAT	TGGCGCAGAC	18360
50	AAATCACATA	CGAGTTGGCT	CTGGTGGTGT	GATGCTGCCG	CACTATCGAC	CTTATAAAAT	18420
	TGCTGAGCAT	TTTAGAATGA	TGGCAGCGTT	ATATCCAAAT	CGTATTGATT	TAGGTATTGG	18480
55							

	TAGTTACGAT GAATCGATTT CGTTATTACG TGATTATCTT ACAATAAAGG ATAAACCAAG	18600
	TGCGCATACG TTAGGTGTCC AACCACACAT TGATCATTTT CCAGAAATGT GGTATTAAAG	18660
5	TAGTAGCGCA ACATCTGCCA AAATAGCTGC CGAACTAGGT ATAGGGCTTT CTGTTGGAAC	18720
	ATTTTGTGCTA CCAGATATAA ATGCGATACA TACAGCGAAG GATAACATTG ATATTTACAA	18780
	AAAACATTTT CAAGCATCAA CGATTAAAAT GGACGCAAAG GTGATGGCAT CTGTATTTGT	18840
10	CATTGTAGCT GATAACGAAG CGGAAGTAGC AGCATTACAA CATGCCTTAG ATGTTTGGTT	18900
	ATTAGGTAAA TTACAATTTG CAGAATTTGA AGATTTTCCT TCAGTAGACA CAGCACAAAA	18960
15	GTATAAGCTT AATGATCGAG ACAAAGAGAT GATTCAAGCA CATCAAGCAC GCATCATTGC	19020
	AGGTACACAA GAAAAGGTTA AAGCACAAAT AGATGATTTC ATTGCTACGT TTGAAGTTGA	19080
	TGAGGTGTTA GTAGCACCGC TTATTCCAGG TATTGAACAG CGTTGTAAAA CATTAAAATT	19140
20	ACTCGCGGAA ATTTATTTGT AGCATTTTAA ATAGAAGAGA AAGGATGAAG ATAAGATGAA	19200
	AAAGTTAGCC AATTATTTAT GGGTAGAAAA AGTAGGAGAT TTGTATGTGT TTAGTATGAC	19260
	ACCTGAATTG CAAGATGATA TTGGGACAGT AGGTTATGTT GAATTCGTAA GTCCAGATGA	19320
25	AGTTAAAGTG GATGATGAAA TTGTGAGTAT CGAAGCATCG AAAACGGTCA TTGATGTGCA	19380
	AACGCCATTG TCAGGAACGA TTATTGAGCG AAATACAAAA GCGGAAGAAG AACCGACAAT	19440
	TTTAACTCT GAAAAACCAG AAGAAAATTG GTTGTTCAAA TTGGATGATG TCGATAAAGA	19500
30	AGCATTCCTA GCATTACCGG AGGCTTAAAT GGAAACGTTA AAATCAAATA AAGCGAGACT	19560
	TGAATATTTA ATCAATGATA TGCATCGAGA GAGAAATGAC AATGACGTAT TGGTAATGCC	19620
	ATCTTCATTT GAAGATTTGT GGAATTATA TCGAGGCTTA GCAAATGTCA GACCGGCATT	19680
35	ACCTGTAAGT GATGAATATT TAGCTGTACA AGATGCTATG TTAAGTGATT TGAATCGTCA	19740
	ACATGTTACG GATTTGAAGG ATTTGAAGCC GATAAAAGGT GACAATATCT TTGTTTGGCA	19800
	AGGTGATATC ACGACGTTAA AAATCGATGC TATTGTTAAT GCTGCAAATA GTCGTTTTCT	19860
40	AGGATGTATG CAAGCTAATC ATGACTGCAT TGATAATATT ATTCATACAA AAGCGGGTGT	19920
	TCAAGTTCGA CTGTATTGTG CAGAGATCAT TCGACAACAA GGGCGCAATG AAGGTGTAGG	19980
45	TAAAGCCAAA ATAACACGTG GATATAATTT GCCAGCAAAG TATATAATTC ATACGGTTGG	20040
	TCCGCAAATA CGTCGATTGC CTGTTTCAAA GATGAATCAG GACTTGTTAG CTAAATGTTA	20100
	TCCTAGCTGT CTTAAATTGG CTGATCAACA TAGTTTAAAT CATGTCGCTT TTTGCTGTAT	20160
50	ATCTACAGGT GTATTTGCTT TTCCTCAAGA TGAAGCAGCA GAAATTGCTG TTCGAACAGT	20220
	AGAAAGCTAT CTCAAAGAAA CAAATTCAAC ATTGAAAGTC GTGTTCAATG TATTTACAGA	20280

55

	CAATGTCTCT GTTAATGGAT GACAAGACAA AGCAGGCTGA AGTATTGCGT ACTGCGATTG	20400
	ATGAAGCAGA TGCATAGTG ATTGGAATTG GTGCAGGCAT GTCTGCATCT GACGGATTTA	20460
5	CATATGTAGG AGAGCGTTTT ACGGAAAATT TCCCAGATTT TATTGAAAAA TATCGCTTCT	20520
	TTGATATGTT GCAAGCGAGT TTACATCCTT ATGGCAGTTG GCAAGAGTAT TGGGCATTTG	20580
	AGAGTCGTTT TATTACATTA AACTATTTAG ATCAACCTGT AGGTCAGTCT TACCTCGCTT	20640
10	TAAAATCCTT GGTGGAAGGT AAACAGTACC ACATTATAAC TACGAATGCA GATAATGCTT	20700
	TCGATGTAGC TGATTATGAT ATGACTCATG TATTTCATAT ACAAGGGGAG TATATACTGC	20760
	AACAGTGTAG CTCAGCATTG TCATGCTCAA ACGTATCGCA ATGATGATTT AATTCGTAAA	20820
15	ATGGTTGTTG CGCAACAAGA TATGCTTATA CCTTGGGAGA TGATTCCAAG ATGTCCAAAA	20880
	TGTGATGCCC CAATGGAAGT GAATAAACGT AAAGCGGAAG TTGGGATGGT TGAAGATGCT	20940
20	GAATTCATG CGCAACTACA TCGTTATAAT GCTTTTCTAG AGCAACATCA AGATGATAAA	21000
	GTGTTGTATT TGGAAATTGG AATTGGTTAT ACTACACCAC AATTTGTGAA GCATCCTTTT	21060
	CAGCGTATGA CACGTAAAAA TGAAAATGCC CTTTATATGA CGATGAATAA AAAGGCATAT	21120
25	CGCATTCGA ATTCAATTCA AGAACGTACC ATACATTTAA CTGAGGATAT CTCAACATTG	21180
	ATTACAGCAG CACTCCGGAA CGACAGCACA ACGAAAAATA ACAACATTGG AGAGACAGAA	21240
	GATGTACTTA ATAGAACCGA TTAGAAATGG AGAATATATT ACTGATGGTG CGATTGCACT	21300
30	CGCTATGCAA GTTTATGTTA ACCAGCATAT CTTTTAGAT GAAGATATTT TATTCCTTA	21360
	TTATTGTGAT CAAAAGTGG AAATTGGACG TTTTCAAAT ACTGCTATAG AAGTGAATCA	21420
	AGATTATATA GATAAACACA GTATTCAAGT AGTTCGCCGA GATACTGGTG GTGGCGCTGT	21480
35	GTATGTTGAT AAAGGTGCCG TTAATATGTG TTGTATTTTA GAACAAGACA CTTCAATTTA	21540
	TGGTGATTTT CAACGATTTT ATCAACCAGC TATAAAGGCG TTGCATACAT TAGGTGCAAC	21600
	AGATGTGGTA CAAAGCGGTA GAAATGATTT AACATTGAAT GGTAAAAAAG TGTCAGGCGC	21660
40	CGCAATGACA TTAATGAATA ATCGTATTTA TGGCGGTTAT TCGCTATTAC TTGATGTTAA	21720
	TTATGAAGCA ATGGATAAAG TGTTAAAGCC TAATCGCAA AAGATTGCAT CGAAAGGGAT	21780
45	TAAATCTGTG CGCGCACGTG TTGGTCATCT TAGAGAAGCA CTGGATGAAA AGTATCGTGA	21840
	TATAACCAAT GAAGAATTTA AAAATTTAAT GGTGACGCAG ATTTTGGGAA TCGATGACAT	21900
	TAAAGAGGCG AAACGATATG AATTAACGGA TGCAGATTGG GAAGCGATTG ATGAATTAGC	21960
50	TGATAAAAAG TATAAAAATT GGGATTGGAA TTATGGCAAG TCACCCAAAT ATGAATACAA	22020
	TCGAAGTGAA AGATTATCTT CAGGTACGGT AGACATAACA ATTTCTGTTG AACAAAATCG	22080

5 AGAAGCATT A CAAGGAACAA AAATGACAAG AGAAGATTTA ACGCATCAGT TAAAGCAATT 22200  
 AGACATCGTT TATTATTTTG GCAATGTTAC GGTAGAAGCA TTAGTGGATA TGATTTTAAG 22260  
 10 TTAATATTGT TATTTTATGT ATGCTGAATC ATTGGAAGTG TTTGCTTGCT CTTGAAAAGG 22320  
 TGACAATAGT GTTTGGTGAA GGTGGAACAT ATGAGTGGAA ATTATTGCCT TTAACATATC 22380  
 AAAGTATGAT ATATATATGG TTTTGTGTTT TAAATGATTG GGTATTTGAA AATAGATGAG 22440  
 15 TTTAATATTT TAAGGAATAT AATGATGTTT ACTTTTATAA TTCATATAGA ATATTAAGCA 22500  
 ATATAAGTCT GTTGATATAT ACAAATATA ATGACTGCTA TAATGAGTAA TCAATAGACA 22560  
 CAAAGAGGAG ATTATGTGAT GAATAATAAA GTATTAGTAA CCGGTGGTAC AGGGTTTGT 22620  
 20 GGCATCGCAA TTATTTACAG ATTATTAGAA CAAGGTTATG ACGTACAAAC GACGATACGT 22680  
 GATTTAAGTA AAGCTGATAA AGTAATTAAC ACAATGCAAG ACAATGGCAT TTCCACAGAG 22740  
 CGATTAATGT TTGTCGAAGC GGATTTATCA CAAGATGAAC ATTGGGATGA AGCAATGAAA 22800  
 GATTGCAAGT ATGTCTTGAG TGAGCATCT CCGGTGTTTT TCGGTAAAAC AGACGATGCA 22860  
 GAAGTGATGG CGAaCTGcAA TTGAAGGTAT ACAACGTATT TTAAGAGCTG CAGAACATGC 22920  
 25 GGGTGTTAAA CGTGTGGTAA TGAAGGTAT ACAACGTATT TTAAGAGCTG CAGAACATGC 22980  
 TAAAAATTCA ATCACAAATG AAAGTCATTG GACAAATGAA GATGAACCAG GCTTATCAGT 23040  
 ATATGAAAAA TCAAAATTGT TAGCTGAAAA GGCAGCGTGG GATTTTGTG AGAATGAAAA 23100  
 30 TACAACAGTA GAATTTGCCA CAATCAATCC AGTTGCAATT TTTGGGCCAT CATTAGATGC 23160  
 ACACGTTTCA GGAAGCTTTC ATTTATTAGA AAATTTATTG AATGGTTCAA TGAAACGTGT 23220  
 ACCGCAAATT CCGTTAAATG TTGTTGATGT GAGAGACGTA GCTGAACTGC ACATTTTGGC 23280  
 35 AATGACAAAT GAACAAGCTA ATGGCAAGCG ATTATTGCG ACGGCTGATG GACmAAATtWA 23340  
 tTTGTTGGGA ATTGcCAAAt TAATTAAAGA AAAGGGCCTG GAAATAGCTC CAAAAGTTCC 23400  
 40 TACTAAAAAA TTACCCAGCT TTATTTTGAG CnAnGnGCC 23439

(2) INFORMATION FOR SEQ ID NO: 39:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4522 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

CCCTTTGAGA GTATATCATC TAGTCAAATT ATGCCTGTCA TTAGAGCGAC TAGCTTTGAT 60

	TATTATGCAG TCGATTTAGG GAAATCATAT CGTCTAATTG ACGAAAGCAT GTTAGAGGAT	180
	TTGAAGTTAA CTGAACAACA AATAAGAGAA ATGTCTCTGT TTAATGTTAG AAAATTGTCA	240
5	AATTCATATA CGACTGATGA AGTAAAAGGT AATATTTTTT ATTTTATTAA CTCAAATGAC	300
	GGGTATGATG CAAGTAGGAT ACTAAATACT GCATTTTAA ATGAAATTGA GGCACAATGT	360
	CAAGGCGAAA TGCTCGTAGC AGTGCCACAC CAAGATGTGT TAATTATTGC AGATATACGC	420
10	AATAAACAG GATATGATGT GATGGCACAT TTAACAATGG AATTTTTCAC TAAAGGTCTA	480
	GTTCCAATTA CATCATTATC CTTTGGATAT AAACAGGGTC ATCTTGAACC GATATTTATT	540
	TTAGGTAAAA ATAATAAACA AAAAAGAGAT CCAAACGTGA TTCAGCGTTT AGAAGCAAAT	600
15	CGTCGTAAAT TTAATAAAGA TAAATAGAAA TAATTGGATA AGGAGTTTGT TCATAATGAA	660
	TTTATTTTAC AATCCTAAAT ATGTAGGAGA TGTCGCATTT TTACAAATTG AACCAGTTGA	720
	AGGTGAATTA AACTACAATA AAAAAGGTAA TGTGTGTGAA ATTACTAATG AAGGTAATGT	780
20	TGTAGGTTAT AATATTTTTG AAATTTCAAA AGATATAACA ATTGAAGAAA AAGGTCATAT	840
	TAAATTAATCT GATGAACTTG TAAATGTATT CCAAAGCGT ATTTCAGAAG CTGGTTTTGA	900
25	TTATAAATTA AATGCTGATC TATCACCGAA ATTGTAGTT GGCTACGTTG AAATAAAGA	960
	CAAACATCCT GATGCAGATA AATTAAGTGT ACTAAATGTA AACGTTGGAA ATGACACATT	1020
	ACAAATTGTA TGTGGCGCGC CTAACGTTGA AGCTGGACAG AAAGTTGTTG TTGCTAAAGT	1080
30	AGGTGCAGTG ATGCCTAGCG GTATGGTAAT TAAAGATGCT GAATTACGTG GTGTTGCCTC	1140
	AAGCGGTATG ATTTGTTCAA TGAAAGAATT GAATTTACCT AATGCACCTG AAGAAAAAGG	1200
	TATTATGGTA TTAAATGACA GCTATGAAAT TGGACAAGCA TTtTTTGAAT AATTAAGGAA	1260
35	GGTAGTGAAA ATATGAGCTG GTTTGATAAA TTATTCGGCG AAGATAATGA TTCAAATGAT	1320
	GACTTGATTG ATAGAAAGAA AAAAAGACGT CAAGAATCAC AAAATATAGA TrACGATCAT	1380
	GACTCATTAC TGCCTCAAAA TAATGATATT TATAGTCGTC CGAGGGGAAA ATTCCGTTTT	1440
40	CCTATGAGCG TAGCTTATGA AAATGAAAAT GTTGAACAAT CTGCAGATAC TATTTCAGAT	1500
	GAAAAAGAAC AATACCATCG AGACTATCGC AAACAAAGCC ACGATTCTCG TTCACAAAAA	1560
45	CGACATCGCC GTAGAAGAAA TCAACAACCT GAAGAACAAA ATTATAGTGA ACAACGTGGG	1620
	AATTCTAAAA TATCACAGCA AAGTATAAAA TATAAAGATC ATTCACATTA CCATACGAAT	1680
	AAGCCAGGTA CATATGTTTC TGCAATTAAT GGTATTGAGA AGGAAACGCA CAAGCCAAAA	1740
50	ACACATAATA TGTATTCTAA TAATACAAAT CATCGTGCTA AAGATTCAAC TCCAGATTAT	1800
	CACAAAGAAA GTTTCAAGAC TTCAGAGGTA CCGTCAGCTA TTTTGGCAC AATGAAACCT	1860

55

	AAACAAAAAT ATGATAAATA TGTAGCTAAG ACGCAAACGT CTCAAAATAA ACAATTAGAA	1980
	CAAGAAAAAC AAAATGATAG TGTGTGCTAAA CAAGGAACTG CATCTAAATC ATCTGATGAA	2040
5	AATGTATCAT CAACAACAAA ATCAATGCCT AATTATTCAA AAGTTGATAA TACTATCAAA	2100
	ATTGAAAATA TTTATGCTTC ACAAATTGTT GAAGAAATTA GACGTGAACG AGAACGTAAA	2160
	GTGCTTCAAA AGCGTCGATT TAAAAAGCG TTGCAACAAA AGCGTGAAGA ACATAAAAAAC	2220
10	GAAGAGCAAG ATGCAATACA ACGTGCAATT GATGAAATGT ATGCTAAACA AGcGGAAcGc	2280
	TATGTTGGTG ATAGTTCATT AAATGATGAT AGTGACTTAA CAGATAATAG TACAGATGCT	2340
	AGTCAGCTTC ATACAAATGG CATAGAGAAT GAACTGTAT CAAATGATGA AAATAAACAA	2400
15	GCGTCAATAC AAAATGAAGA CACTAATGAC ACTCATGTAG ATGAAAGTCC ATACAATTAT	2460
	GAGGAAGTTA GTTGAaTCA AGTATCGACA ACAAACAAT TGTCAGATGA TGAAGTTACG	2520
	GTTTCGAATG TAACGTCTCA ACATCAATCA GCACTACAAC ATAACGTTGA AGTAAATGAT	2580
20	AAAGATGAAC TAAAAATCA ATCCAGATTA ATTGCTGATT CAGAAGAAGA TGGAGCAACG	2640
	aATAAAGAAG AATATTcAGk AAGTCAAATC GATGATGCAG AATTTTATGA ATTAAATGAT	2700
25	ACAGAAGTAG ATGAGGATAC TACTTCAAAT ATCGAAGATA ATACCAATAG AAACGCGTCT	2760
	GAAATGCATG TAGACGCTCC TAAACGCAA GAGTACGCAG TAACTGAATC TCAAGTAAAT	2820
	AATATCGATA AAACGGTTGA TAATGAAATT GAATTAGCAC CGCGTCATAA AAAAGATGAC	2880
30	CAACAAACT TAAGTGTCAA CTCATTGAAA ACGAATGATG TGAATGATAA TCATGTTGTG	2940
	GAAGATTCAA GCATGAATGA AATAGAAAAG AATAACGCAG AAATTACAGA AAATGTGCAA	3000
	AACGAAGCAG CTGAAAGTGA ACAAATGTC GAAGAGAAAA CTATTGAAAA CGTAAATCCA	3060
35	AAGAAACAGA CTGAAAAGGT TTCAACTTTA AGTAAAAGAC CATTTAATGT TGTcATGACG	3120
	CCATcTGATA AAAAGCGTAT GATGGATCGT AAAAAGCATT CAAAAGTCAA TGTGCCTGAA	3180
	TTAAAGCCTG TACAAAGTAA GCAAGCTGTG AGTGAAAGAA TGCCTGCGAG TCAAGCCACA	3240
40	CCATCATCAA GATCTGATTC ACAAGAGTCA AATACAAATG CATATAAAAC AAATAATATG	3300
	ACATCAAACA ATGTTGaGAA CAATCAACTT ATTGGTCATG CAGAAACAGA AAATGATTAT	3360
	CAAAATGCAC AACAATATTC AGAGCAGAAA CCTTCTGTTG aTTCAACTCA AACGGAAATA	3420
45	TTTGAAGAAA GTCAAGATGA TAATCAATTG GAAAATGAGC AAGTTGATCA ATCAACTTCG	3480
	TCTTCAGTTT CAGAAGTAAG CGACATAACT GAAGAAAGCG AAGAAACAAC ACATCCAAAC	3540
	AATACTAGTG GACAACAAGA TAATGATGAT CAACAAAAAG ATTTACAGTC ATCATTTTCA	3600
50	AATAAAAATG AAGATACAGC TAATGAAAAT AGACCTCGGA CGAACCAACA AGATGTTGCA	3660

55



CCAAGTGTTT CATTACTAGA AGAACCACAA GTTATTGAGT CGGACGAGGA CTGGATTACA 3780  
 GATAAAAAGA AAGAACTGAA TGACGCATTA TTTTACTTTA ATGTACCTGC AGAAGTACAA 3840  
 5 GATGTAAGTG AAGGTCCAAG TGTTACAAGA TTTGAATTAT CAGTTGAAAA AGGTGTTAAA 3900  
 GTTTCAGAA TTACGGCATT ACAAGATGAC ATTAATGATGG CATTGGCAGC GAAAGATATT 3960  
 CGTATAGAAG CGCCTATTCC AGGAACTAGT CGTGTGGTA TTGAAGTCC GAACCAAAAT 4020  
 10 CCAACGACAG TCAACTTACG TTCTATTATT GAATCTCCaA GTTTTAAAAA TGCTGAATCT 4080  
 AAATTAACAG TTGCGATGGG GTATAGAATT AATAATGAAC CATTACTTAT GGATATTGCT 4140  
 AAAACGCCAC ACGCACTAAT TGCAGGTGCA ACTGGATCAG GGAAATCAGT TTGTATCAAT 4200  
 15 AGTATTTTGA TGTCTTTACT ATATAAAAAT CATCCTGAGG AATTAAGATT ATTACTTATC 4260  
 GATCCAAAAA TGTTGAATT AGCTCCTTAT AATGGTTTGC CACATTTAGT TGCACCGGTA 4320  
 ATTACAGATG TCAAAGCAGC TACACAGAGT TTAAATGGG CCGTAGAAGA AATGGAACGA 4380  
 20 CGTTATAAGT TATTTGCACA TTACCCATGT ACGTAnTATA ACAGCATTTA ACnAAAAAGC 4440  
 CCCATATGAT GAAAGAATGn CAAAAATTGT CATTGTaATT GATGAGTTGG CTGATTTAAT 4500  
 25 GATGATGGTC CGCAAGAAGT TG 4522

(2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 751 base pairs  
 30 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

TCAAGTTTAC GGATACGTAT ATATTTTGCA TGACATTTAG TGCAATAATA TTCATAATTT 60  
 GCCCGTTGTT GATAGCTTTC AATGCTGTGA CAAAATCTAG GCGCTCCAAC CTGTTGGCTC 120  
 40 AATCGTTTAA AATCTTGATC TTTATGTTGA TAACCTTTAC CAGCAATATG CAAGTGATAA 180  
 TGACACAATT CGTGCAGTAT AATTTTTACA ACAGCATCTT CTCCATAATG CTCATATTGT 240  
 TTTGGATTAA TTTCAATATC ATGGGACTTT AAAAGATAAC GTCCGCCTGT TGTACGTAAC 300  
 45 CTTTTATTAA AATATGCACA ATGTCGAAAC GTACGTCCAA ATTTTTCTTC CGAAAGATTC 360  
 TCAACCATTG GCTGAAGTTT GTCATTATTC ATGTGGATCA ATCATCGTTA ATGATACTTT 420  
 GTCTTTATTT TTGTCAATAC TGTAATCCA AACGTCAACG ATATCACCAA CACTGACAAT 480  
 50 ATCCATTGGA TTTTTTACGA ACTTCTTAGA AAGTTTCGAA ACATGGACAA GTCCATCTTG 540

55

TTTCATTCTT TCTTGTAAT CTTCAATTGA TAGCACATCG GATTTAAGGA TTGGTGTTC 660  
 AACTCGTCC CTTGGATCTC GATTAGGTGC GTTCAAGGAT TTAATAATAT CCTCTAATGT 720  
 5 AGGTACACCG ACTTGTAATT CAATCGCCAG T 751

(2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1076 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

TCTCCAGCTT TAACTTGATC TGGCACTTTA ACAATTGTCT GATCCATACA TACGCGACCA 60  
 20 ATAACCTCGC ATTGATGACC ATTTACATTT ACAAAGCTAC CTTGCATTAT GCGTAAATGG 120  
 CCATCTGCAT ATCCAATAGG TAACAATGCT ATTGTAGTTG GGTCAGTAGC TGTATAAGTT 180  
 GCACCATAAC TTACAGACTC ACCCGCTTGT AGCGTCTTTG TTTGAACTAC ATTAGCAATT 240  
 25 AATTGCACAC TTGGTTTAAG GTGTACTTTA ACTTTTGTCT GTACATACTC TGATGGATAA 300  
 TATCCATAAA GGGAAATTCC TGGTCTTATT GCATTACAGA ATTGGCAATC CATTAATAGA 360  
 GAGCCTGCTG AGTCTGACA ATGTATATAT TCAGGTTTAA TTGCTTCATT GACCATATCT 420  
 30 TTAAACGTT GATATTGTTT AGTTGTCATA TCTCCTGGTT CGTCAGCACA GGCAAAGTGT 480  
 GTAAACACGC CTTCAAATAC AAGTTGCTCA TATGTTGAA TGATTTCAAT CACTTCTTGA 540  
 TACGTTTTAG TATCTTTAAT ACCTAAACGT CCCATTCCTG TATCTAATTT AATGTGCAAC 600  
 35 CATACTTTT TCTCTTGCTC ACCAGAAATG TTTTAAATTG CTTCTTTCAA CCACTGTTTA 660  
 GACGGAACCG TTAAGGCAAC TCGGTGTTGT ATCGCTTTAT CAATATCTTT AGCTGGTAAC 720  
 ACACCTAAGA CTAAAATTTT AGCAGTAATC CCATGCATTC TAAGTTCTAT CGCTTCATCT 780  
 40 AACGTTGCTA CAGCAAAAAA TGTGGCGCCA TTTTCCATTA AATGACGTGC TACTTTAACA 840  
 CTACCTAGTC CATAGGCATT GGCTTTAACG ACAGCCATCA CTGTTTTATT TGGATGCAAT 900  
 GTACTGAATA CTTTGAAATT TGATGCAACA GCGTTTAAAT CTACATTCAT ATACGCAGAT 960  
 45 CTATAATATT TATCCGACAT ATTACTTCCT CCTGTAATTC CCACACGTTT TAAAACTAGA 1020  
 TCTTAATTAT CATTGTATAA CAAATTTAAA ATGCTGACTT TTCTAAAACA ACTTGG 1076

(2) INFORMATION FOR SEQ ID NO: 42:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2930 base pairs

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

	TGACCACAAT GCCCAATACA ACCATCCCAT GGTAAGCCA AGAGATGAGT CAATAAAGCG	60
10	TGTTGAATAA GAGCTGAATG AACCTGATAC TGGATAAAAT GTTGCCAACT CTCCAATTGA	120
	TGACATTAAG AAATATAGCA TGACACCAAT AACAAGATAA GCGAGTATAG CGCCTCCAGG	180
	ACCAGCTTGA GAAATGATAT TACCAGTAGC TACAAATAGA CCAGTCCCAA TTGCACCACC	240
15	TATAGCAATC ATGGAAATGT GTCTTGAGTT AAGACTACGG TTCATTTTAT TATCTTCCAT	300
	ATTIAGTCTC CCATCTATTT AAATATACCC ATTATTGTAA GCTTTTAAAG TGTACTATTC	360
	AATAACTATT TAGTACTGTA AAGCGAAAAA ATTAAATTT TCTGATTTTT TAATCATCTT	420
20	GAGCATGTTT AATTGTAATT TTGATGGGGT TAAATTATAA TATGTATTAA ATTATAATTA	480
	TnATAAATTG TGGAGGGaTG ACTATGTCAC AACAAGACAA AAAGTAACT GGTGTTTTTG	540
	GGCATCCAGT ATCAGACCGA GAAAATAGTA TGACAGCAGG GCCTAGGGGA CCTCTTTTAA	600
25	TGCAAGATAT TTACTTTTTA GAGCAAATGT CTCAATTGA TAGAGAAGTA ATACCAGAAC	660
	GTCGAATGCA TGCCAAAGGT TCTGGTGCAT TTGGGACATT TACTGTAACT AAAGATATAA	720
	CAAAATATAC GAATGCTAAA AtATTCTCTG AAATAGGTAA GCAAACCGAA ATGTTTGCCC	780
30	GTTTCTCTAC TGTAGCAGGA GAACGTGGTG CTGCTGATGC GGAcGTGACA TTCGAGGATT	840
	TGCGTTAAAG TTCTACACTG AAGAAGGGAA CTGGGaTTTA GTAGGGAATA ACACACCAGT	900
	ATTCTTCTTT AGAGATCCAA AGTTATTTGT TAGTTTAAAT CGTGCGGTGA AACGAGATCC	960
35	TAGAACAAAT ATGAGAGATG CACAAAATAA CTGGGATTTT TGGaCGGGTt TCCAGAAGCA	1020
	TTGCACCAAG TAACGATCTT AATGTCAGAT AGAGGGATTC CTAAAGATTT ACGTCATATG	1080
40	CATGGGTTTC GTTCTCACAC ATACTCTATG TATAATGATT CTGGTGAACG TGTTTGGGTT	1140
	AAATTCCATT TTAGAACGCA ACAAGGTATT GAAAACTTAA CTGATGAAGA AGCTGCTGAA	1200
	ATTATAGCTA CAGATCGTGA TTCATCTCAA CGCGATTTAT TCGAAGCCAT TGAAAAAGGT	1260
45	GATTATCCAA AATGGACAAT GTATATTCAA GTAATGACTG AGGAACAAGC TAAAAACCAT	1320
	AAAGATAATC CATTTGATTT AACAAAAGTA TGGTATCAG ATGAGTATCC TCTAATTGAA	1380
	GTTGGAGAGT TTGAATTAAA TAGAAATCCA GATAATTACT TTATGGATGT TGAACAAGCT	1440
50	GCGTTTGAC CAACTAATAT TATTCCAGGA TTAGATTTTT CTCCAGACAA AATGCTGCAA	1500
	GGGCGTTTAT TCTCATATGG CGATGCGCAA AGATATCGAT TAGGAGTTAA TCATTGGCAG	1560

55

GGTCAAATGC GCGTAGTTGA CAATAACCAA GGTGGAGGAA CACATTATTA TCCAAATAAC 1680  
 CATGGTAAAT TTGATTCTCA ACCTGAATAT AAAAAGCCAC CATTCCCAAC TGATGGATAC 1740  
 5 GGCTATGAAT ATAATCAACG TCAAGATGAT GATAATTATT TTGAACAACC AGGTAAATTG 1800  
 TTTAGATTAC AATCAGAGGA CGCTAAAGAA AGAATTTTAA CAAATACAGC AAATGCAATG 1860  
 GAAGGCGTAA CGGATGATGT TAAACGACGT CATATTCGTC ATTGTTACAA AGCTGACCCA 1920  
 10 GAATATGGTA AAGGTGTTGC AAAAGCATTG GGTATTGATA TAAATTCTAT TGATCTTGAA 1980  
 ACTGAAAATG ATGAAACATA CGAAAACTTT GAAAAATAAA TTTGATATGT AGTTTCTATA 2040  
 TTGCGTAGTT GAGCAGTTTA TGATATCATA ATAAATCGTA AAGATTCCTA ACAAGAGAGG 2100  
 15 GTGTTTAAACG TGCGCGTAAA CGTAACATTA GCATGCACAG AATGTGGCGA TCGTAACTAT 2160  
 ATCACTACTA AAAATAAACG TAATAATCCT GAGCGTATTG AAATGAAAAA ATATTGCCCA 2220  
 AGATTAAACA AATATACGTT ACATCGTGAA ACTAAGTAAT TCTTATCATT CAAATACGAC 2280  
 20 GATTTGAAAA TAAAGCGGGC TTACCTATTA TATTGGGGAG CTCGCTTTTT TATGAAATTT 2340  
 TTGTGAAGAG TGATTAATGG ATTGAGTTTC ATCGGTAGAA CAATATATGA TTATATTAGT 2400  
 TGTTACTTTA TTAaaaTTTG AGAATATTTA TAGAAGCAAA TAGATTACTG ATTTTATAAA 2460  
 25 GTCACTTTGT TAGCGAATGC TTGAAAGAGT ATTTAATATA GTAGAATTTA AAATTTCAAA 2520  
 GCGGAATTTA ATAAGTACGA AGTAGTTCGT GGTATGTTTT ATAAATGTTC GATAATACAC 2580  
 30 TTTAATCTTA AATATGATGG TTTAGAAAAT GATTTAACAA AGAAATGAaa CTTTACTGTT 2640  
 GAATTATGTG AGGATTGTGT TATTATATAA ATCGTAATAA TTACGATTTG ATAAAAAGTG 2700  
 AGGTAACATAT ATATGGCTAA GAAATCTAAA ATAGCAAAAG AGAGAAAAAG AGAAGAGTTA 2760  
 35 GTAAATAAAT ATTACGAATT ACGTAAAGAG TTAAAAGCAA AAGGTGATTA CGAAGCGTTA 2820  
 AGAAaATTAC CAAGAGATTC ATCACCTACA CGTTTAACTA GAAGATGTAA AGTAACTGGA 2880  
 AGACCTAGAG GTGTATTACG TAAATTTGAA ATGTCTCGTA TTGCGTTTAG 2930

40 (2) INFORMATION FOR SEQ ID NO: 43:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3606 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 45 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

50 CTTCTTGCCA TGGCTCTCTT TATTAAAAA TGCTTCCAAC TTGTCCATTT GATTGTTTCT 60  
 55

	TTATAAAAAA CTAATTTTAC AAATGCTTTT GCGTTCTTAC AAAAAATGCA TTTGACTATT	180
	ATTATAATAA GCGTATAATT GTCGCATATT ATTTTTTGTA TTTTGGCAA TAACGAAGGA	240
5	GTATTTATGA ATAAAGACAA GCAATTGCAC AACGACAAAA TCAATCTATC CCAATTAGTC	300
	TTATTAGGGT TAGGCTCTTT AATAGGATCT GGTGGCTAT TTGGTGGTG GGAAGCATCA	360
	TCAATAGCTG GACCAGCAGC AATCATATCA TGGGTTCTTG GATTCCTAGT CATTGAAC	420
10	ATTGCCTATA ACTACATTGA AATCGGCACA ATGTTTCCTC AATCAGGTGG CATGAGTAAC	480
	TATGCCAGT ATACACATGG CTCATTATTA GGCTTTATTG CTGCTGGGC GAATTGGGTG	540
	TCTTTGGTGA CAATAATACC TATCGAAGCT GTGTCAGCTG TTCAATATAT GAGTTCTTGG	600
15	CCGTGGCATT GGGCGAAACC AATGAGATAT TTAATGGAAA ATGGCTCTAT TAGCACATAC	660
	GGATTGCTAG CTGTATATCT CATCATGTGTT ATTTTTTCAT TATTAACTA TTGGTCCGTA	720
	AAACTTTTAA CATCATTTAC GAGTTTAATT TCTGTATTTA AATTAGGCGT ACCCATGTTA	780
20	ACCATCATCA TGGTGATGCT ATCAGGATTC GACACTTCAA ATTACGGCCA TTCGGCAAGC	840
	ACATTTATGC CTTACGGAAG TGCACCGATT TTTGCTGCAA CAACAGCATC AGGGATTATT	900
25	TTTTCATTCA ATTCATTCCA GACAATTATT AATATGGGTT CAGAAATTAA AAATCCTGAA	960
	AAAAATATCG CAAGAGGCAT CGCTATCTCA CTGTCAATCA GTGCAGTGTT GTACATCATT	1020
	TTACAAAGTA CGTTTATCAC TTCTATGCCT CAATCAATGT TACAACATAG TGGATGGAAT	1080
30	GGCATCAACT TCAATTCACC ATTTGCTGAT TTAGCTATCT TATTAGGAAT TAATTGGCTC	1140
	GCAATTTTAC TATACATTGA AGCTTTTGTA TCACCATTCG GTACTGGCGT GTCATTTGTC	1200
	GCCGTTACAG GTCGAGTTTT ACGAGCAATG GAGAAAAATG GACATATCCC TAAATTTCTT	1260
35	GGGAAGATGA ATGAAAAATA TCATATCCCA CGTGTAGCAA TCATCTTTAA TGCCATCATT	1320
	AGTATGATTA TGGTTACATT ATTTAGAGAT TGGGGTACGC TAGCAGCAGT TATTTCTACT	1380
	GCAACTTTAG TAGCCTATTT AACTGGCCCA ACGACAGTGA TTGCATTAAG AAAAAATGGA	1440
40	CCAACAATGA CTCGTCCATT TAGAGCAAAA ATTTTAAAAG TAATGGCACC ATTATCATT	1500
	GTATTAGCTT CATTAGCTAT ATATTGGGCA ATGTGGCCAA CAACGGCTGA AGTTATTTTA	1560
	ATCATTATAC TTGGATTACC AATCTACTTC TTCTATGAAT ATCGTATGAA TTGGCGTAAT	1620
45	ACAAAGAAAC AAATTGGTGG TAGCTTATGG ATTATTGTAT ATTTAATCGT GCTATCAATA	1680
	CTGTCATTTA TAGGAAGCAA AGAATTTAAA GGCTTAAATA TGATTCATA TCCATTTGAC	1740
	TTTATCGTTA TTATTATTGT GGCATTATC TTCTATTACA TCGGTACAAC GAGTTCAATT	1800
50	GAAAGCGTCT ATTTCCGTCG CGCAACACGA ATCAATACGA AGATGCGTGA GTCACTAAAT	1860

55

	CACACACATT AACCAACCAT TGATTTC AAC ATCTTGGTTG GTTTTTTATT TTGAAAATCG	1980
	GTTATAAATA ACTAACATAA CAAGATGATG ATCAGGCTGG GACATAAATC AATGTTCTAT	2040
5	GCTCTACGAA gTTATATTGG CAGTAGTTGA CTGAACGAAA ATGCGCTTGT AACAAAGCTTT	2100
	TTTCGATTCT AGTCAGGGGC CCCAACACAG AGAATTTCTGA AAAGAAATTC TACAGGCAAT	2160
	GCAAGTTGGG GTGGGACGAC GATAAAGAAA TACTTTTTCT ATAGAAATTA GTATYtCTTA	2220
10	TGCAAGAGTT TTA CTCTATGT ATTCATATTT TTAAGTACAC ATTAGCTGTG GCTAATGTAT	2280
	AAGAACCCT ACATAATAAA TCATTTGTGG CTCTTATCA TTTCTGTCCC ACTCCCGTAG	2340
	AAGTACATCA TATAATGCTG AAAATGGTTT GAGTTAAAAC AGATATCAAG CTCGTCTGAT	2400
15	TCACTCACAA AATTGTCTTG TTATACTTGT CACCTATCAT CTATAGACCG TGGTATGATT	2460
	AAATTGGGGA TGATAAGGA GGTAAATAAA TATGAAGATT AATACTACAG GTGGTCAAAT	2520
	TCATGGTATT ACACAAGATG GTTTAGATAT CTTCTTAGGC ATTCCTTATG CAGAACCACC	2580
20	AGTTCATGAC AATCGCTTTA AACATTCTAC GTTAAAAACA CAATGGTCAG AGCCAATTGA	2640
	TGCAACTGAA ATACAACCCA TCCCACCGCA ACCAGACAAC AAATTAGAAG ATTTTTTCTC	2700
	CTCACAATCT ACAACTTTTA CTGAACATGA AGACTGTTTA TATCTAAATA TTTGGAAACA	2760
25	ACATAATGAT CAGACGAAGA AACCTGTCAT CATTTATTTT TATGGTGGTA GTTTTGAAAA	2820
	TGGTCATGGT ACAGCCGAAC TCTATCAACC GGCACATTTA GTACAAAATA ACGACATTAT	2880
30	CGTTATTACA TGCAATTATC GTTTAGGCGC ATTAGGATAT TTAGACTGGT CATATTTTAA	2940
	TAAAGATTTT CATTCCAATA ATGGCCTTTC AGATCAAATC AATGTCATAA AATGGGTGCA	3000
	TCAATTTATT GAATCCTTCG GTGGCGACGC TAATAACATT ACTTTAATGG GTCAGTCTGC	3060
35	AGGCAGTATG AGCATTTTGA CTTTACTTAA AATACCTGAC ATTGAGCCAT ACTTCCATAA	3120
	AGTCTTCTA CTAAGTGGCG CACTACGATT AGACACCCTT GAGAGTGAC GCAATAAAGC	3180
	ACAACATTTT CAAAAATGA TGCTCGATTA TTTAGATACA GATGATGTTA CATCATTATC	3240
40	GACAAATGAT ATTCTTATGC TGATGGCGAA gcTAAACAA TCTCGAGGAC CTTCTAAAGG	3300
	GCTTGATTTA ATATATGCGC CTATTAAAAC AGATTATATA CAAAATAATT ATCCAACAAC	3360
	GAAACCAATT TTTGCATGTT ATACAAAAGA TGAAGGCGAT ATTTATATTA CTAGTGAACA	3420
45	GAAAAAATTA TCGCCGCAAC GCTTTATCGA CATTATGGAA TTAAATGATA TTCCTTTAAA	3480
	ATACGAAGAT GTTCAGACGG CGAAGcAACA ATCTTTAGCG ATTACACATT GTTATTTCaA	3540
	ACAGCCGATG aAGCAATTTT TACmACmACT CAATATACmA GATTCCAACC GCACCAACTA	3600
50	TGGCTT	3606

55

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 15109 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

10 GAAATTAAAA AAGCAATTGG TACAAGATGC AACAGTGTCA TTGTTTGATG AATTTGATAA 60  
 AAAATTATAC ACTTACGGCG ATAAGTGGG TCGTGGTGA GAAGTATTAT ATCAAGCATT 120  
 TGGTTTGAAG ATGCAACsAG AACAACAAAA GTTAACTGCA AAAGCAGGT GGGCTGAAGT 180  
 15 GAAACAAGAA GAAATTGAAA AATATGCTGG TGATTACATT GTGAGTACAA GTGAAGGTAA 240  
 ACCTACACCA GGATACGAAT CAACAAACAT GTGGaAGAAT TTGAAGCTA CTAAAGAAGG 300  
 ACATATTGTT AAAGTTGATG CTGGTACATA CTGGTACAAC GATCCTTATA CATTAGATTT 360  
 20 CATGCGTAAA GATTTAAAAG AAAAAATTAAT TAAAGCTGCA AAATAATTCA GCTATATAAG 420  
 TTAGTGAAAT GAGAGTCTGA AACATATCAA TCTTTTGATA TTGTATTAGG CTCTTATTTT 480  
 TATAGCTAGA AAGTTAGATA TTTGTATTTT TTTAAATAAT AAGTGCCGTT GTTATCGTTC 540  
 25 AATTTAATTA ATGATAGATT AGTATTATTA TAGCTAAAGT AGTATACCTG AGAAAAATAGC 600  
 TCAATGTATC TCTTTATTAA TAAGTTATAT CATAATTATT TTAGTGATA CTTTATGGAA 660  
 30 GGGATATCAG GGAATGGCTT TCAATTAAAG AAGAGGTTTA AAAGGATTAC AACAGAATGT 720  
 TATGATTTTG TAGAAAGATA TATAACAACG TTTTATAAAA ACATAATATT GTTAATGGAA 780  
 AATGAAATGT AAGGGGGATT TCGAGTGAAT AAGAAAGTTT ATTTTAACCA CGATGGTGGT 840  
 35 GTAGATGATT TAGTATCTCT ATTTTATTA TTACAAATGG AAAACGTTCA ATTGATAGGG 900  
 GTCAGTACAA TTGGTGCTGA TTGTTATTTA GAGCCATCTT TGAGCGCATC AGTAAAAATT 960  
 ATTAATCGTT TTTCAAATGA AGATATTCAA GTTGCGCCAT CATATGAACG AGGAAAAAAT 1020  
 40 CCATTTCTTA AAGAATGGCG TATGCATGCC TTTTATATGG ACGCATTGCC AATTTTAAAT 1080  
 GAGCCAGTCA AACATGTTGC TTCAAATGTG AGCGACAAAG AAGCCTTTGA AGACATTATT 1140  
 CAAACTTTAA AGAGACAATC AGAAAAAGTA ACATTATTAT TTACAGGCCC GCTTACAGAT 1200  
 45 TTAGCAAAAG CACTACAAAA AGATTATCT ATCGTTCAGT ATATAGAAAA ATTAGTTTGG 1260  
 ATGGGTGGCA CCTTTTACC AAAAGGAAAT GTTGAAGAAC CTGAGCATGA TGGTCTGCA 1320  
 GAATGGAATG CATATTGGGA TCCAGAAGCG GTTAAATTTG TTTTGTAGAG CGATATAGAG 1380  
 50 ATTGATATGG TTGCTTTAGA AAGTACGAAT CAAGTACCGC TAACGTTAGA TGTTAGACAA 1440

55

	GTACCACCAT TAACACACTT TATAACAAAT TCTACTTACT TTTTATGGGA TGTTTTAACG	1560
	ACTGCTTATA TTGGTAACAA GGACTTGGTT CATTCAATTG AGAAAAAGT CGATGTAATA	1620
5	AGTTATGGAC CAAGTCAAGG TAAGACATTT GAGTGTAAG ATGGGCGCAA AATTAATGTC	1680
	ATAAATCATG TAGATAACAA CGCATTTTTT GATTATATAA CTGCACTTGC TAAAAAGTA	1740
	AATTAACAGC TGTGTAGAAT AATTAAGGTT TTAATTTATA TAGAACAACT TATTGTAAAC	1800
10	TTTTCATTTT TTAAGTTTAA CAATGGTGCT ATAATAATGG TCATGAAATA CGAAAGGAAG	1860
	TAAAAAATGA CAACAAAACA GTTAGTATAT ACAGCTTTAA TGACAGCGAT TATCGCTATT	1920
	TTAGGATTGG TACCGGTAAT TCCACTACCA TTTCTTCAG TACCAATTGT ACTTCAAAC	1980
15	ATTGGTATTT TCTTAGCAGG TGCGATTTTA GGACGTAAAT ATGGCACATT AAGTGTTATC	2040
	GTCTTTTTAT TATTAGTAGT TGCTGGCTTG CCATTGTTAT CAGGTGGTCG CGGTGGCATC	2100
	GGTGTAATCG CAGGTCCTTC AGCAGGGTTT TTACTATTAT ATCCAGTTGT AGCATTTCATG	2160
20	ATTGGGGCGA TTCGAGATAG ATTCATCAAT GAAATTAATT TCTGGATTTT ATTCGTTGGT	2220
	ATTTTAGTTT TTGGTGTTAT AGCATTAGAT GTTATTGGTA CATTGATTAT GGGCATGATT	2280
	ATTAACATAC CATTTACGAA AGCTATTTCA ATTTCAATTAG CTTATTTGCC TGGTGATATA	2340
25	TTAAAAGCAA TTGTAGCAAG TTTGATTGGT ACAGCTTTAC TTAATCACTC GCAGTTTCGT	2400
	CAAAATTATG GAATAAATA ATCATATTTA AGATAGTAAA GTAATTGAAT AAGTTGCTTT	2460
30	GAAATTTATA AAAGTGAAAG GAGTAGGTGT CAATGGCTAG TATAAGTATG TCAGATATAT	2520
	ATTGTAACGG CACTATATTT GAAATGACG ACGAGCAGTT GATTATTTA ACGCCTTCTT	2580
	TTCCACAACG ATACACAAGT AACACATGGA TATATAAAAA GACGCCTACC CAAGAGCGAT	2640
35	GGCTGAAAGA CTTAGAACGT CAACATCAAT TACATACAAA TCAAGGTTCA AATCATTATG	2700
	CGTTTAGTTT CCCGAAAAAT GAACAACCTG ATAATCATTG GATGGCTATG TTAAAGATA	2760
	TGAATTTTGA ACTAGGTATT ATGGAATTGT ATGCCATAGA AAGTGATGCG CTTGCCAATT	2820
40	TGCCGCGTAA CTCTGACGTT GAAATTGCCA TCGTTGACGA GTCGCATATA GATGCCTATT	2880
	TAAAAGTTGC ATATCAGTTT AGTTTGCCAT TTGGAAGAAG CTATGCAGAT GCACATGAAG	2940
	AAATGGTAAG GGAACATTAT CAAAAAGATG TGATTAAACG CTTAGTAGCT TATTTAAATA	3000
45	ATGAACCTAT TGGCGTTGTA GATGTCATTG AAAGTGAAAA TTACATTGAA TTAGATGGAT	3060
	TTGGTGATT AGAACAATTT CGGCACCAAG GAATTGGATC TACAATTCAA TCGTTGATAG	3120
	GTGAATACGC CATATCAAAA AATCACAAAC CAATCATATT AGTTGCAGAT GGTGAAGATA	3180
50	CAGCAAAAGA TATGTATGCA AAGCAAGGTT ATGTCTATCA ATCGTTTTGT TATCAAATAT	3240

55



	TAAGCTGGTT TCGAGTAGAA ATCAACTTAC TGCTTTTTAA ATTGTTTTGA GCTACTTATA	3360
	CTTATAAAAA TAGTGCCTTT AAATTGTTGA TTCATGTAGA ATATCGTTCA TTATGACACA	3420
5	CTATAATGAA TATGTTATTG TTCAGAATCA ATGATACGTT CTGGATGACT GTATATATTA	3480
	AAGCCACCAT TTCGAATAAA TCCAACGGCC GTAATATTTA GGTCATTAGC TAAGGTTACA	3540
	GCAAGCGTTG TCGGAGCTGA TTTAGATAAA ATGACGCCAA CACCAATTTT TCGGGCTTTA	3600
10	ATTAAAAATT CTGATGAAAT ACGTCCACTA AAAATTAATA CTTTATCTCG GACAGTAATA	3660
	TGTCGCTGAA TACAAAAATC ATATAATTTA TCTAGAGCGT TATGTCTACC AATGTCTTGT	3720
15	CGATGTACAA AAAATGTCAA ACCATCGCTT ATAGCAGCAT TATGTAAGCC ACCTGTTTCT	3780
	TGGTAAATAT GACTTGCACT TTGTAATCGA GTCATCATGT TAATAATTG CATTGGAGTT	3840
	AAAGTGATTT TAGACATAGA TGTTTTAGCG ATAGCAGCAT CATTTTGAAA ATAAAACTCA	3900
20	CGACTCTTTC CGCAACAAGA TGCAATCATT CGTTTTGTGG AATATTGAAA GCGATCGCCT	3960
	AAATCTTTAT TAAGTTCAAC ATGGGCAAAA CCTTTACTAT CATCAATCAG TACAGATTTT	4020
	AATTCATCTC GCTTTAAAT GGCACCTTCC GAAGCCAGAA ATCCAATGAC TAACTCCTCA	4080
25	AGGTTTGTG GACTGCATAT AACAGTCGCA AATTCTTCAC CATTCAACCAT AATTGTAAGT	4140
	GGAAATCTG TCACATATTG ATCTGTTGTA TTGAATAATT TTCCATCTTC ATATCTAACA	4200
	ATTGGTTGAC CTAAAGATAC ATCTTTGTTT ATTATCTAAC CCCTTTAATT AGCTTAAACT	4260
30	TTATTTTAAA GCAATTTGCT TAAAAATTTA ACATATTTGC TTAAGTTTGA AATTGATTG	4320
	ATAAAAAATTA ATAGCGAGCA ATCTGTTTGA TTTAAATTGA ATTCGAGAAT ATACATACTA	4380
35	GGGCATCAAT TAATAAATAT CAATCTTATG CAAATTTGAC AATTGTTTGA ATCAATATAT	4440
	AAACAGGCAA CGGTTCTTTT CAAATATAAT AGTAAGTGTA TAATGAAAAT GTAAATATTA	4500
	TTAAAAATGG GGGTTCATC AATGAAATTG AAACGTTTAT TTGCTGTTGT GATTGCAATG	4560
40	CTTTTAGTAT TAGCTGGTTG CTCTAATTCT AACGATAATA ATGAAAGTAA AAAAGATGAC	4620
	GCAGACAATG GTAAGAAACA AGAGATTCAA GTTGCAGCGG CAGCAAGTTT AACAGATGTA	4680
	ACCAAGAAAT TAGCTTCAGA ATTTAAAAAA GAGCATAAAA ATGCTGATAT TAAATTTAAC	4740
45	TATGGTGGAT CAGGGGCATT AAGAAAACAA ATTGAATCAG GCGCACCTGT TGACGTATTT	4800
	ATGTCTGCAA ATACTAAAGA TGTAGATGCA TTAAAAGACA AGAATAAAGC GCATGATACA	4860
	TATAAATATG CGAAAAATAG TCTAGTATTA ATTGGTGATA AAGATTCAAA TTCACTTCA	4920
50	GTAAAAGACT TAAAAGACAA TGATAAATTA GCATTAGGTG AAGTGAAAAC TGTACCAGCA	4980
	GGAAAATATG CGAAACAGTA TTTAGATAAC AATAACTTAT TTAAAGAAGT CGAAAGTAAA	5040
55		

	CAAGGTTTTG TGTATAAAAC TGA	CTTATAT AAACAAAATA AAAAAATTGA TACTGTAAAA	5160
	GTAATTAAAG AAGTAGAACT TAAGAAGCCA ATCACATACG AAGCTGGTGC TACATCAGAT		5220
5	AGTAAATTAG CAAAAGAGTG GATGGAATTC TTTAAATCAG ATAAAGCTAA AGAAATACTA		5280
	AAAGAATACC ACTTTGCAGC ATAAGGAGTT GTAATCCATG CCTGACTTAA CACCTTTTTG		5340
	GATATCAATA CGAGTTGCTG TAATCAGTAC GATTATTGTA ACGGTTTTAG GTATTTTAT		5400
10	ATCTAAATGG TTGTATCGTC GTAAGGGTTC GTGGGTAA GTATTGGAAA GTTTATTGAT		5460
	ATTACCTATT GTTTGCCGC CAACGGTATT AGGTTTTATT CTATTAATCA TCTTCTCGCC		5520
	AAGAGGACCA ATCGGTCAAT TCTTTGCGAA TGTACTACAT TTACCTGTAG TGTTCACCTT		5580
15	GACAGGTGCT GTGATAGCAT CTGTCATTGT TAGTTTTCCA CTAATGTATC AACATACTGT		5640
	GCAAGGCTTC AGAGGTATAG ACACGAAAT GATTAATACA GCTAGAACGA TGGGAGCAAG		5700
20	TGAAACGAAA ATTTCTCTCA AATTAATTTT ACCATTAGCT AAACGCTCTA TTTTAGCAGG		5760
	TATAATGATG AGTTTTGCTC GTGCATTAGG TGAGTTTGGT GCTACATTAA TGGTTGCAGG		5820
	ATATATTCCA AATAAACGA ATACACTACC TTTAGAAATA TACTTCTTAG TGGAAACAAGG		5880
25	TAGAGAAAAT GAAGCGTGGT TATGGGTATT AGTGCTAGTC GCATTCTCTA TTGTGGTTAT		5940
	ATCTACAATT AATTTATTGA ATAAAGATAA ATATAAGGAG GTCGACTAGA TGCTTAAAT		6000
	CAATGTGAAA TATCAATTAA AGAACCTTT AATTCGCATC AATATAGATG ATACTGAACC		6060
30	AAAAATTTAT GCAGTTCGTG GTCCATCTGG CATTGGTAAA ACTACTGTTT TAAATATGAT		6120
	TGCCGGATTA CGTAAAGCAG ATGAAGCTAT TATCGAAGTG AATGGGCAAT TACTTACTGA		6180
	TACGGCAAAA AACGTGAATG TTAAATTC AACAACGACGT ATTGGATATC TGTTC AAGA		6240
35	CTACCAATTG TTTCTAATA TGACGGTCTA TAAAAATATT ACTTTTATGG CTGAACCATC		6300
	TGAAACACATC GATCAATTAA TTCAAACCTT AAACATTGAT CATTTGATGA AACAATATCC		6360
40	TATGACATTG TCAGGTGGAG AGGCACAACG TGTAGCACTT GCACGTGCAC TTAGCACAA		6420
	ACCAGATTTA ATTTTATTAG ATGAACCTTT TTCTAGTTTG GATGATACTA CAAAAGATGA		6480
	GAGTATTACA TTAGTTAAAC GTATTTTCAA CGAATGGCAA ATACCAATCA TATTTGTGAC		6540
45	ACATTCAAAC TATGAAGCAG AACAAATGGC TCATGAAATT ATTACAATTG GGTAATCATT		6600
	TATTTGCCAT TAAAGAGTTT AGAACGTATT TAAAATTGTA GAAGTGAATG CTTCTATCAG		6660
	CATTTTAATG ATGTTTTAAA CTCTTTTTTA GGGGCAGTTT TTTTGAGAGA CATTGACGCG		6720
50	CGTCATATAA TGAAAGTAAT GATAAAAAGA AAGGATAACT TAATGTGAGT CAAGAACGTT		6780
	ATTCAAGGCA AATTTTATTT AAACAAATAG GTGAAATAGG TCAAAGCAAA ATAAATCAAA		6840

55

	GAGCAGGCAT TGCCAAACTA ATCATTGTTG ATAGAGATTA TATTGAATTT AGTAATTTAC	6960
	AAAGACAAAC ATTGTTTACT GAAGAAGATG CTTTGAAAAT GATGCCTAAG GTGGTTGCAG	7020
5	CTAAAAAGCA TTTGCTAGCG TTACGTAGTG ATGTTGATAT TGATGATTAT ATTGCCCATG	7080
	TGGATTATTA TTTTGTGGAA ACACATGGAC AGGACGTTGA CGTTATTATT GATGCAACCG	7140
	ATAACTTTGA AACACGACAA CTGATTAATG ATTTTGCATA TAAATATCGT ATACCTTGA	7200
10	TTTATGGTGG TGTGTACAG AGTACATATA CAGAAGCTGC ATTTATACCT GGTAAAAACAC	7260
	CTTGCTTTAA CTGTTTGGTA CCACAATTGC CAGCATTAAT TTTAACATGT GATACAGTAG	7320
	GGGTCATTCA ACCTGCCGTG ACGATGGCAA CAAGTTTACA ATTAAGAGAT GCGATGAAAG	7380
15	TATTAACGGA ACAACCAATT GACACAAAAA TAACTTATGG CGATATTTGG GAAGGTAGTC	7440
	ATTATTCATT TGGTTTCAGT AAAATGCAAC GTTCAGACTG TACAACTTGT GGAGATGTAC	7500
20	CAAGTTATCC GTATTTAAAC AAGAATGAAC AACGTTATGC AACATTGTGT GGTAGAGACA	7560
	CTGTACAGTA TGAATGCA TCAATTACAC ACGACATTCT TGTTCAATTT TTAACAAC	7620
	ATCAGTTAAA TTATCGCAGT AATTCGTATA TGGTTATGTT TGAATTTAAA GGACACCGCA	7680
25	TTGTTGCTTT TAAAGGTGGA AGGTTTTTAA TACATGGCAT GACACGCACA TCAGATGCCA	7740
	CACATCTAAT GAATTTATTG TTTGGATAAA AAAAGATAAG ACAAAGGAG TGTAATATTA	7800
	TGGGCGAACA TCAAAACGTT AAATTGAATC GTACAGTTAA AGCAGCCGTA CTAACGGTAT	7860
30	CAGATACTAG AGACTTTGAT ACAGATAAAG GTGGTCAATG CGTGCGCCAA CTATTACAAG	7920
	CAGATGACGT TGAAGTGAGT GACGCACATT ATACAATTGT GAAAGATGAA AAAGTAGCCA	7980
	TCACGACGCA GGTGAAGAAG TGGTTAGAAG AAGATATTGA TGTCATCATT ACGACTGGTG	8040
35	GAACAGGTAT TGCACAACGT GATGTGACGA TTGAAGCAGT AAAACCACTT TTAATAAAG	8100
	AGATAGAAGG CTTTGGGGAA TTGTTTAGAT ATTTGAGTTA TGTTGAAGAT GTTGGCACGC	8160
40	GTGCATTATT GTCTCGTGCT GTAGCAGGTA CAGTTAATAA TAAATTGATA TTTTCGATTG	8220
	CAGGATCAAC AGGCGCAGTT AAATTAGCAT TAGAAAAGCT CATTAAACCA GAATTAAATC	8280
	ATCTGATTCA TGAGCTTACA AAATAATTTA TTGATTGAT TGGCGTTGAA AATCTCCAGA	8340
45	TTTACCGCCA GACTTGCTTT CAAGGTAGGT TTCGCCAATA ATCATACCTT TATCAACTGC	8400
	TTTCGTCATG TCGTAAATGG TTAAAGCCGT TGCTGATGCA GCGGTTAAAG CTTCCATTTC	8460
	AACACCGGTT TTGCCAGTTG TAGAGACAGT TGTTTGAATG TTTAAAGTAT AAAGGGGTGC	8520
50	ATTTGTTTCA TCCCAGCTGA AGTGAACATC TATGCCAGTC AATGGTAATG GATGGCACAT	8580
	CGGAATAAGT GTTGATGTAT TTTTGGCAGC CATAATACCA GCGATTTGAG CAGTGTTCAA	8640
55		

	AATGCTTGAA TGAGCGACAG CAGTTCITTT TGTAATTTGT TTGTCTGATA CATCGACCAT	8760
	TTTGGCGTGG CCTTGTTGAT TAATATGAGT AAACCTCAGTC ATTTTACCCC TCCTAGTGCA	8820
5	TCTAGTATAT CATGAAAAA TAAAAGTTTT GGACATGATT TTTAATGGTA GTAGAAAAA	8880
	GAAACCCAAT CCCAGTTAAA GAAGCAATTC AACGTATCGT TAATCAGCAG AGTTCAATGC	8940
10	CGGCAATTAC GGTAGCACTT GAAAAAGTC TAAATCATAT CTTAGCAGAA GATATTGTAG	9000
	CTACTTATGA TATACCAAGG TTTGATAAAT CACCTTATGA TGGTTTTCGA ATTTCGAGTG	9060
	TTGATTCACA AGGGGCAAGT GGTGAGAATC GCATTGAGTT TAAAGTGATT GATCATATTG	9120
15	GTGCAGGTTT AGTTTCTGAT AAATTAGTTG GGGATCACGA AGCGGTGCGT ATTATGACTG	9180
	GAGCACAAAT ACCTAATGGC GCAGATGCTG TTGTTATGTT TGAACAAACG ATTGAAGTAG	9240
	AAGATACATT TACAATTCGT AAACCATTTT CAAAAATGA AAATATATCT TAAAAGGTG	9300
20	AAGAAACAAA GACAGGCGAT GTTGTCTTAA AAAAAGGACA AGTAATTAAT CCAGGGGCTA	9360
	TCGCGGTCCT TGCAACATAT GGCTATGCAG AGGTTAAAGT TATTAAGCAA CCGAGTGTCG	9420
	CTGTTATTGC AACAGGAAGC GAATTATTAG ATGTTAATGA TGTATTAGAA GATGGGAAAA	9480
25	TTCTGTAATC TAATGGCCCA ATGATTTCGT CCTTAGCAGA AAAATTAGGT CTGGAAGTTG	9540
	GTATTTACAA AACACAAAAA GATGATTTAG ATAGTGGCAT CCAAGTCGTT AAAGAAGCTA	9600
	TGGAAAAACA TGATATCGTT ATTACAACGG GCGGAGTTTC TGTGGAGAT TTGACTATT	9660
30	TACCTGAGAT TTATAAGGCT GTAAAGGCGG AAGTGTATT TAATAAAGTA GCAATGCGTC	9720
	CTGGTAGCGT AACACCGGTT GCATTTGTAG ATGGaAAGTA TTTGTTTGGa TTATCTGGAA	9780
	ATCCATCAGC TTGTTTTACA GGATTTGAAC TATTTGTGAA nCCAGCTGTT AAACATATGT	9840
35	GTGGCGCACT AGAAGTCTTC CCGCAAATAA TTAAAGCAAC ATTAATGGAA GATTTTACCA	9900
	AGGCAAACCC ATTCACACGA TTTATACGTG CTAAAGCAAC GTTAACAAGT GCTGGAGCTA	9960
40	CTGTAGTACC TTCAGGATTC AATAAATCAG GTGCGGTTGT AGCGATTGCA CATGCTAACT	10020
	GTATGGTCAT GTTACCAGGA GGGTCACGTG GTTTTAAAGC GGGGCATACA GTAGATATTA	10080
	TATTGACTGA ATCTGACGCT GCTGAAGAGG AACTTCTTTT ATGATTTTAC AAATTGTAGG	10140
45	TTACAAAAAG TCTGGTAAGA CAACATTGAT GAGGCATATT GTCTCTTTCT TAAAGTCACA	10200
	TGGTTATACA GTTGCTACTA TTAAACATCA TGGGCATGGT AAGGAAGATA TTCAATTACA	10260
	GGATTCAGAC GTCGATCACA TGAAGCATTT TGAAGCGGGG GCAGATCAAA GTATTGTACA	10320
50	AGGTTTTCOA TATCAGCAAA CTGTAACACG TGTAGATAAT CAAAATCTTA CTCAAATTAT	10380
	TGAAAAATCT GTTACAATTG ACACCAATAT CGTATTAGTT GAAGGCTTTA AAAATGCTGA	10440

55

	GAATGTTTGT TATAGCATTA ATGTAAGGGA GCATGAAGAT TTTACAGCAT TTGAGCAATG	10560
	GTTATTAAAT AAAATTAAAA ATGATTGTGA TACACAATTA ACATAGAGGA TTGAAATGAA	10620
5	TGAAACAATT TGAAATCGTG ACAGAACCGA TACAAACAGA ACAATATCGT GAATTCACCTA	10680
	TAAATGAATA TCAAGGTGCA GTAGTTGTTT TTACCGGTCA TGTTCCGCGAA TGGACTAAAG	10740
	GCGTCAAAAC GGAATATTTA GAATATGAAG CGTATATTCC AATGGCTGAA AAGAAATTGG	10800
10	CACAAATTGG AGATGAAATA AATGAAAAAT GGCCTGGAAC GATAACGAGT ATTGTTTCATA	10860
	GAATAGGGCC ATTACAAATT TCAGATATCG CTGTATTAAAT TGCGGTTTCT TCACCGCATC	10920
	GTAAAGATGC CTATCGAGCA AATGAATATG CAATTGAGCG TATAAAAGAA ATTGTTCCGA	10980
15	TTTGGAAGAA AGAAATTTGG GAAGATGGTT CAAAATGGCA AGGGCATCAA AAAGGGAATT	11040
	ATGAAGAAGC AAAGAGGGAG GAATAAGAGA GATGAAGGTA CTTTACTTCG CAGAAATTAA	11100
20	AGATATATTA CAAAAGCAC AGGAAGATAT TGTGCTTGAA CAAGCATTGA CTGTACAACA	11160
	ATTTGAAGAT TTATTGTTTG AACGTTATCC GCAAATCAAT AATAAAAAGT TTCAAGTTGC	11220
	TGTAAATGAG GAATTTGTAC AAAAATCGGA TTTTCATTCAA CCTAATGATA CTGTTGCATT	11280
25	AATTCCACCG GTTAGTGGAG GTTAAGGGAG CATGAAAGCA ATAATTCTTG CAGGTGGTCA	11340
	TTCAGTGCGA TTTGGTAAGC CCAAAGCTTT TGCGGAAGTG AACGGTGAGA CCTTTTATAG	11400
	TAGAGTAATT AAGACATTAG AATCAACAAA TATGTTCAAT GAAATTATTA TTAGTACAAA	11460
30	TGCGCAATTG GCAACGCAAT TTAAATATCC AAATGTTGTT ATAGATGATG AGAATCATAA	11520
	TGATAAAGGT CCATTAGCAG GAATTTATAC AATCATGAAG CAACATCCTG AAGAAGAATT	11580
	GTTTTTTGTC GTTTCTGTTG ATACACCAAT GATTACTGGT AAAGCTGTAA GCACGTTGTA	11640
35	TCAGTTTTTA GTTTCTCATC TTATTGAAAA TCATTTAGAT GTCGCAGCTT TTAAAGAAGA	11700
	TGGACGTTTT ATTCCAACAA TTGCATTTTA TAGTCCGAAT GCATTAGGCG CTATAACTAA	11760
	AGCACTACAT TCTGATAATT ACAGTTTTAA AAATGTATAT CATGAATTAT CAACGGATTA	11820
40	TTTGATGTA AGGGATGTAG ATGCGCCCTC ATATTGGTAC AAAAATATAA ATTATCAGCA	11880
	TGATTTGGAC GCTTTAATTC AAAAATTGTA AGCTGTTAGG AGGTCCACAA ATGGTAGAAC	11940
45	AAATAAAAGA TAAACTAGGA CGTCCCATCC GTGACTTACG GTTATCTGTG ACAGATCGGT	12000
	GTAACTTTAG GTGTGATTAT TGCATGCCTA AAGAGGTATT TGGAGATGAT TTCGTATTTT	12060
	TACCTAAAAA TGAACTTTTA ACGTTTGATG AAATGGCTAG AATCGCTAAG GTATATGCAG	12120
50	AATTAGGTGT AAAAAAATA CGCATTACAG GTGGAGAACC ATTGATGCGA CGGGATTTAG	12180
	ATGTACTTAT AGCTAAATTA AATCAAATCG ATGGTATTGA AGATATTGGT TTGACTACAA	12240

55

	ATGTCAGTTT GGATGCTATT GATGATACGC TATTTCAATC AATCAATAAT CGTAATATTA	12360
	AAGCGACTAC GATTTTAGAA CAAATTGATT ACGCGACGTC TATTGGTTTG AATGTAAAAG	12420
5	TAAATGTTGT TATACAAAAA GGTATTAAACG ATGATCAAAT CATACCAATG CTTGAATATT	12480
	TTAAAGATAA ACATATAGAG ATTCGATTTA TAGAATTTAT GGATGTTGGT AATGATAATG	12540
	GATGGGATTT CAGTAAAGTT GTAACATAAG ATGAAATGCT TACAATGATA GAGCAGCACT	12600
10	TTGAAATCGA TCCTGTAGAA CAAAAATATT TTGGGGAAGT AGCAAAATAT TATCGCCATA	12660
	AGGATAATGG TGTTCAATTG GGTTCGATTA CAAGTGTTC ACAATCATT TGTCTACAT	12720
	GTACACGCGC AAGGCTGTCA TCAGATGGGA AGTTTTACGG ATGTTTATTT GCAACTGTCTG	12780
15	ATGGATTAA CGTTAAAGCG TTTATTCGTT CTGGCGTGAC CGACGAAGAA TTTAAAGAAC	12840
	AATTTAAAGC TTTATGGCAA ATAAGAGATG ATCGATATTC AGATGAGAGA ACTGCTCAAA	12900
20	CAGTTGCCAA TCGTCAACGT AAAAAGATAA ACATGAATTA TATTGGTGGT TAATGTGTAG	12960
	GGACCACTAC ATATTAAATC ATTAGAGATG TTTAATATT TCTGTCTTAC TCCCTAAAT	13020
	ACAATATTAT TTATTAAAGT AAAAACGGTC ATATCTATGC CAGATTTAAT AGAAATGATC	13080
25	GTTTTTAAAG TTTTACAAG TTGGCGGGGC CCCAACACAG AAGCTGACAG AAAGTCAGCT	13140
	TACAATAATG TGCAAGTTGG CGGGCCCCCA ACATAGAGAA TTTCAAAAAG AAATTCTACA	13200
	GACAATGCAA GTTGGGGAAC GGGGCCCCAA CACAGAAGGT GACGAAAAGT CAGCATACAA	13260
30	TAATGTGCAA GTTGGCGGGG CCCCAACATA GAGAATTTC AAAGAAATTC TACAGACAAT	13320
	GCAAGTTGGG GATCAACGAA ATAAATTTTA TGAGAATATC ATTTCTATCC CACTCTTAAG	13380
	AATCACTACA TAATAAATCT TTAGTGGTTC TTAAACATTG ATGTCACACT CCATGCCATT	13440
35	GAGTTGTAAT ATATCTTTTT TAGGTATAAA TGTTGTCGAA TAAACAACAA GTTGTCCAA	13500
	AGATATAAAT CTAAACAAGA TATAGCCAGC AATTTAATAT TTGTAATAGA TAAAATGCTA	13560
	AGTTTGATAT ATAATAAATT TAAGTAATTG TATAATAATA TGAATTACAA ACATCTAAGA	13620
40	AGAAACATAG GAGGCATCAT ATTATGAGTA ATAAAGTTCA ACGTTTTATA GAAGCAGAAA	13680
	GGGAGTTAAG TCAGTTAAAG CACTGGTTAA AAACAACACA TAAGATTTCA ATTGAAGAAT	13740
45	TTGTAGTCCT TTTTAAAGTG TATGAAGCTG AAAAGATTAG CGGTAAAGAA TTGAGGGATm	13800
	CATTACATTT TGAAATGCTA TGGGATACAA GTAAATCGA TGTGATTATC CGTAAATCT	13860
	ATAAAAAAGA GCTTATTTCT AAATTGCGTT CTGAAACGGA TGAAAGACAA GTATTCTATT	13920
50	TCTATAGTAC TTCTCAAAAG AAATTGTTAG ATAAATTTAC TAAAGAAATA GAAGTGTAA	13980
	CGGTTACAAA CTAAAACTT aAAAagcaTG CCAATCTCTA TTCATCATAA TTGCGTCTTG	14040

55

GTTCATGGCA TTTCTAGTTA CATGACGTCC ATGAATTAAG AAGTAAACAA GCATAGTAAT 14160  
 GATTGCTAAA GCGGCCATAA AGCCGAAGAT TTCACTATAT GAAAACATAT GAGTAAATAA 14220  
 5 CCCAAGGAAT GATGGACCGA AGCCGACACC TGCATCTAGA CCAACGTAAA AAGTAGATGT 14280  
 CCGGATACCA TATTTAATCG GGGGTGAGAC TTTTATCGCA ATAGATTGCA TTGCAGATGA 14340  
 TAAATTTCCA TACCGTAAAC CTAGGCAAGC ACCAGCAAGT AATATTAACC AGCTTTGATA 14400  
 10 GCTTGAAATT AAGCATACAA ATGAAAGGAA AAGCATGATA AATGCTGGGT AGACAATAAT 14460  
 ATTTTCATTT TTATCATCCA TCAATCTACC AGCAATAGGT CTAGTAATTA ACGATGCTAT 14520  
 AGCATAGCAA ATAAAGAAAT AGCTTGCTGC AGTGACTAGG TGTCGCTCTA AAGCAAATGC 14580  
 15 TTGTAAATAA GTTAGGATGG ACGCATAGGT AACGCCAATT AAAAGCATAA TTACAGCAAC 14640  
 AGGAATGGCC TCTTTTGCAA TAAATTGATG AATACTAAAT CTTGGTTTAT CAATGACATT 14700  
 AGTTTCAGTT TTGTTATTTG TTACTTCGAA ATCAACTTTT ATAAATAATG AGATAATGAG 14760  
 20 TCCGAGTATG CCTAATATGA CACAAATAAT AACAGTAAG TCAATTGCGT ATTTTGTAAT 14820  
 AAGTAACATG CCTAGAAATG GGCCAATCGC TGTACCTAAT ACTAACTTA AGGAAAATAA 14880  
 ACTGATGCCT TCACTTTTTC TATTAACAGG GGTAACGTAT GCCGCAATAG TACCTGTTGC 14940  
 25 AGTTGTCACA ACTGCAGTTG CGATACCGTT TATGAGACGT ACAAAGATTA AAAAAGCTAA 15000  
 AGATCCATCA ATAAATATAA GTAATTGCGT GATAATTAAA GCAATTAAAC CAATAAATAA 15060  
 30 TAATCGTTTA GGTCCrATTT sATTTACAAA TTTACCTGTA GCAAATCGA 15109

## (2) INFORMATION FOR SEQ ID NO: 45:

## (i) SEQUENCE CHARACTERISTICS:

- 35 (A) LENGTH: 9072 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

40 GAGAGTCAAT GGCAAGAAGA ATATAAATAT TTGAGAGCGT TAATCTTTAA TGAAACAGAA 60  
 TTAGAGGAAG CGTATAAATG GATGCATCCT TGTTACACGT TGAATAATAA AAATGTAGTA 120  
 45 CTTATCCATG GCTTCAAAAA TTATGTTGCA CTATTATTTT ATAAAGGTGC CATTTTGGAG 180  
 GATAAATATC ATACACTCAT TCAACAGACT GAAAAGGTGC AAGCAGCTCG TCAGTTACGA 240  
 TTTGAAAATT TAACAGAGAT TCAAGCACGT ACCGAAGAAA TTAAATATTA TCTAGCCGAA 300  
 50 GCAATTAAAG CTGAAAAAGC TGGTAAAAAA GTTGAAATGA AGAAAACAGA GGAATATGTT 360

55

	AAATTAACGC CAGGCAGACA ACATCAATAT ATATATCATA TTGGACAAGC TAAACGCagT	480
	GgAACAAAGAC AAAAGCGTGT TGAAAAGTAT ATTAACCAAA TACTAGAAGG TAAAGGGATG	540
5	CATGATAAGT AATTAATGAG TAAAGCATAC CGGTTATACA ACAACATACA AGATGACACG	600
	AAACAACCAA TGGCTCATGC TGTGGTTGT TTTTTAGGT GTGTCTGTCA TGGGCAACAC	660
	TTTGACGTTG GAATTCGTT ACAGGCTTGG GAGTAGAAAA TGTTAGCAA AGGCAAGGGT	720
10	GTCTACAATG AATGATGAAG ATATTAAAAT ATAAGGATGA CTTTGTGAGT GGCGGATGGG	780
	CGGTTGTCCG TCTGTAACAA TGGATGCGTG TGCATTATTA CAAAAATTCG ACTTTTGTA	840
	TAATATTICA CATTTTCGAC ACTTTTTTGC TATAAACAA CCAATTGAGC GATAATAAAT	900
15	TCGCTTTTAA AAAATATGAG TTATCTATTT AGTTGCCAAA GATAAAATAA TAATGTTTAA	960
	TAACATCATA TAGAGTATGT TAGTTTTAA TGTCGAATAT ACGAATGTGc AAACAAAGTA	1020
20	ATCGGTAGAA ATTCAACATA CATAGCGCCG TTTACTGTTA AGTATTCACA TTACAGATGA	1080
	AAAATATAAA ATTCTACATA ATCAAGACCA TGATGTGTAC TTGTTTAACT TATGACTCTA	1140
	TTTGTTTAAC AATTGCGATA ATGGTCTTTT TATTTTATGC GTATCATTCTG TCATATTTTT	1200
25	TATGAGGAAG GAGAAATGAT TATGTTAAGT ATTAAGCATT TAACGAAAAT TTATTCTGGT	1260
	AATAAAAAGG CAGTAGATGA CATCTCTTTA GATATTCAAT CTGGGGAATT TATCGCATT	1320
	ATTGGAACCA GTGGAAGTGG CAAAACGACT GCTTTAAGAA TGATAAACCG TATGATTGAA	1380
30	GCGACAGAAG GACAAATTGA AATTGATGGT AAAGATGTTC GGAGTATGAA TCCTGTGAA	1440
	TTGCGTAGAA ATATTGGCTA TGTATTCAA CAAATTGGCT TAATGCCTCA TATGACGATT	1500
	AAAGAGAATA TTGTGTTGGT ACCCAAATTG TTGAAATGGA CTAAAGAGGA AAAGGATAAA	1560
35	CGTGCAAAGG AATTAATTAA ACTTGTGGAT TTACCGAGT CATTTTTAGA GCGTTATCCA	1620
	GCAGAACTAT CAGGTGGGCA ACAACAACGT ATCGGTGTTG TAAGAGCACT TGCGGCCGAA	1680
	CAAGATATTA TTTTAATGGA TGAACCTTTT GGTGCATTGG ATCCTATTAC GAGAGATACG	1740
40	TTACAAGATT TAGTTAAAC GTTACAACGA AAATTAGGCA AGACGTTTAT CTTTGTAACA	1800
	CATGATATGG ATGAAGCGAT TAAATTAGCA GACAAAATTT GTATTATGTC AGAAGGTAAG	1860
45	GTGGTGCAAT TTGATACGCC AGACAATATT TTAAGACATC CCGCAAATGA TTTGTACGT	1920
	GATTTTATAG GACAAAATAG ACTGATTCAA GACCGTCCCA ATGACAAGAC TGTAAGAGT	1980
	GTAATGATTA AACCAATCAC GATACAAGCA GAAGCAACAC TGAATGACGC CGTTCATATT	2040
50	ATGAGACAAA AACGTGTTGA TACTATTTTT GTAGTAGATA GTAATAACCA TTTACTAGGT	2100
	TTCTTAGACA TTGAAGATAT AAATCAGGGT ATACGTGGAC ACAAAGTTT ACGAGACACC	2160

55



	ATTTTAAAAA GAAACGTTAG GAATGTACCT GTCGTAGATG ATCAACAGCG TTTAGTAGGA	2280
	CTGATTACGC GTGCCAATGT TGTTGATATT GTATATGACA CGATTGGGG CGATAGTGAG	2340
5	GATACAGTGC AAACAGAACA TGTGGGGGAA GACACTGCGT CCTCAAAAGT GCATGAGCAA	2400
	CACACTACTA ATGTCAAAGT ACGTGACATA GGAGATGATA AATCATGATT GAGTTCCTAC	2460
	ATGAACATGG TGGACAGTTG ATGTCGAAAA CACTGGAACA TTTCTATATT TCTATAGTGG	2520
10	CATTATTACT TGCCATCATT GTTGACAGTAC CTATAGGCAT TTTATTATCA AAAACAAAGC	2580
	GAACTGCCAA TATTGTATTA ACTGTGGCAG GTGTCTTACA AACTATTCCA AACTAGCTG	2640
	TACTTGCTAT TATGATACCG ATTTTGGTG TTGGTAAAAC GCCTGCAATT GTAGCGCTAT	2700
15	TTATTTATGT ATTATTACCT ATTTTAAATA ACACGGTACT CGGTGTTCAA AATATTGATA	2760
	GCAACATTAA AGAAGCTGGA AAAAGTATGG GAATGACACA ATTTCAATTG ATGAAGGATG	2820
20	TTGAATTGCC GTTAGCAATT CCGCTTATCA TTGGTGGCAT TCGTTTGTC TCTGTGTATG	2880
	TAATTAGTTG GGCTACACTT GCAAGTTATG TAGGTGCGGG TGGATTAGGT GATTTCAATT	2940
	TCAATGGTTT AAATTTATAT GATCCACTGA TGATTGTAAC TGCAACGGTA CTCGTTACTG	3000
25	CACTAGCATT AGGTGTTGAT GCCTTATTAG CTTTAGTTGA AAAATGGGTA GTTCCCAAAG	3060
	GCTTAAAAGT ATCTGGATAA TTAGGAGGCT AAGATAATGA AGAAAATTAA ATATATACTT	3120
	GTCGTGTTTG TCTTATCGCT TACCGTATTA TCTGGATGTA GTTTGCCCGG ACTAGGTAGT	3180
30	AAGAGCACGA AAAATGATGT CAAAATTACA GCATTATCAA CAAGCGAATC GCAAATTATT	3240
	TCACATATGT TACGGTTGTT AATAGAGCAT GATACACACG GTAAGATAAA GCCAACATTA	3300
	GTAAATAATT TAGGGTCAAG TACGATTCAA CATAATGCCT TAATTAATGG GGATGCTAAT	3360
35	ATATCAGGTG TTAGATATAA TGGCACAGAT TTAACGGGAG CTTTGAAGGA AGCACCAATT	3420
	AAAAATCCTA AGAAAGCAAT GATAGCAACA CAACAAGGAT TTAAAAAGAA ATTTGATCAA	3480
	ACGTTTTTTG ATTCGTATGG TTTTGC GAAT ACGTATGCAT TCATGGTAAC GAAGGAAACC	3540
40	GCTAAAAAAT ATCATTTAGA GACAGTTTCA GATTTAGCAA AGCATAGTAA AGATTACGT	3600
	TTAGGTATGG ATAGTTCATG GATGAATCGT AAAGGCGATG GCTATGAAGG ATTTAAAAAA	3660
45	GAGTATGGTT TTGACTTTGG TACAGTGAGA CCAATGCAAA TAGGTCTAGT CTACGACGCA	3720
	TTAAACTCAG AGAAGTTAGA CGTTGCATTA GGTATTCTA CAGATGGTCG AATTGCGGCG	3780
	TATGATTTGA AAGTACTTAA AGATGATAAA CAATTTTCC CACCTTATGC TCGGAGTGCT	3840
50	GTTGCAACAA ATGAATTATT ACGGCAACAC CCAGAACTTA AAACGACGAT TAATAAGTTG	3900
	ACAGGAAAGA TTTGACTTC AGAGATGCAA CGCTTGAATT ATGAAGCGGA TGGTAAAGGT	3960

55

	AAAGGTGGTC ATAAGTAATG GAAGGTAATT TATTACAGCA ATTATTCAAT TATTATGTTA	4080
	CGAACTTTGG TTATCTATGG GATTTATTTT TCAAACACTT ATTAATGTCT GTCTATGGTG	4140
5	TGCTGTTTGC AgCTTTAATT GGTATTCCAT TGGGAATCTT GCTTGCaAGA TACACAAAAC	4200
	TTTCTGGATT TGTAATTACA ATTGCAAATA TAATTCAAAC AGTTCCAGTC ATTGCAATGT	4260
	TAGCTATTTT AATGTTAGTC ATGGGCTTAG GTTCAGAAAC AGTAGTTTTA ACAGTGTTTT	4320
10	TATATGCGTT ACTTCCAATT ATAAAAACA CTTATACTGG TATAGCTAGT GTTGATGCGA	4380
	ATATTAAAGGA TGCTGGCAAA GGTATGGGAA TGACACGCAA TCAAGTGCTA CGAATGATTG	4440
	AATTACCGTT ATCTGTTTCG GTTATTATCG GTGGCATTCTG TATTGCCTTG GTTGTTCGCA	4500
15	TAGGTGTTGT TGCCGTTGGA TCATTTATAG GAGCACCTAC GCTTGGTGAC ATTGTGATTC	4560
	GTGGTACAAA TGCCACGGAT GGCACAACGT TTATTTTAGC AGGTGCGATT CCGATTGCTA	4620
	TCATTGCAAT CGTCATTGAT GTACTATTAA GATTTTATAGA AAAACGATTA GACCCAACAA	4680
20	CACGACATCG TAAAAATCAA TCTAATCATC GGCCGCAAAG TATTAAATATG TAATAGTAGA	4740
	AGATGTTTAT AATTTAGCGA TTTCGTTTCA TGATTTATAA AAAATGAGGC TACTCAAGGA	4800
25	GCTCAAATAA TCTTTGAGTA GCCTTTTAT AGGTTGTGTT TGTATGCGTT TACACTAAAA	4860
	TAGCAATTAT TATCATGAAA GTTTTTGGAT AAAAAGCGTT AATTATTGTA AAAATACTAA	4920
	AAAATGAGAT GTTTTATTTA TAATTTCTG CAAATTTATG ATATTGTTTC TTAATATATC	4980
30	ATATTAAAAA TTTGTTTTTC TTAAACATAG GAGGCTTATC TAATTCATGG ACACATCAAA	5040
	ACAATTTAGA GGTGACAACC GATTGCTTTT GGGTATCGTT TTAGGGGTTA TTACCTTTTG	5100
	GCTATTCGCG CAGTCACTTG TTAATCTTGT TGTCCCATTA CAATCAACAT ATAGTAGTGA	5160
35	CGTTGGAACG ATAAATATCG CTGTTAGCTT ATCTGCCTTA TTTGCTGGTT TGTTTATCGT	5220
	AGGTGCTGGT GATGTTGCTG ATAAATTGG TCGCGTCAAA ATTACTTATG TAGGATTGAT	5280
	ATTAAATGTT GTAGGTTTCT TACTCATCAT CATTACACCT TTGCCAGCAT TTTTAATTAT	5340
40	AGGTAGAATA ATTCAAGGTT TGTCTGCAGC ATGTATTATG CCATCAACAC TTGCTATTAT	5400
	TAACGAATAT TATATTGGTA CAAGAAGACA ACGTGCCTTA AGCTATTGGT CTATTGGTTC	5460
45	TTGGGGTGGT AGTGGTATTT GTACGTTGTT TGGTGGCTTA ATGGCTACAT ATATAGGTTG	5520
	GCGTTCAATA TTTGTTGTTT CAATTCTATT AACATTATTA GCAATGTACT TAATCAAACA	5580
	TGCACCTGAG ACTAAAGCAG AACCAATCAA AGGTATGAAA GCAGAAGCTA AAAAGTTTGA	5640
50	CGTTATTGGT TTAGTCATTT TAGTAGTGAC GATGTTAAGT TTAAATGTAA TCATCACACA	5700
	GACGTCTCAT TTTGGTTTAG TTTCACCGTT AATTCTAGGT TTAATTGTTG TGTTTATCTG	5760

55

	AATTTTAAAA AATAGAGGAT ACAGTGGTGC AACTATTTCA AACTTCTTAT TAAATGGTGT	5880
	AGCAGGTGGT GCACTTATCG TTATTAACAC GTATTATCAA CAACAATTAG GATTTAATTC	5940
5	TTCGCAAACG GGTATATTTT CATTAACGTA TTTAATAACA GTGTTGTCAA TGATTCGTGT	6000
	AGGTGAAAAG ATTTTATCTC AACATGGTCC GAAGCGCCCA CTATTACTAG GAAGTGGCTT	6060
	TACASTGATT GGGTTAATCT TATTGTCGTT AACATTTTTA CCAGAAGTGT GGTATATCAT	6120
10	ATCTAGTATA GTTGATATT TATTGTTTGG TACTGGTTTA GGATTATATG CTACACCATC	6180
	AACTGATACA GCAGTTGCTA GTGCGCCAGA TGATAAGTCG GGTGTTGCTT CAGGTGTGTA	6240
	TAAAATGGCG TCATCATTAG GAAATGCATT TGGAGTAGCA GTATCTGGTA CGGTTTATAC	6300
15	TGTGTTAGCA GCTAATTTAA ATTTGAACTT AGGTGGTTTC ACAGGTATGA TGTTTAATGC	6360
	CTTGCTAGCA ATTGTTGCAT TTTTAGTCAT TTTACTATTA GTTCCTAAAA ATCAAACGAA	6420
	TTTGTAAGAC TGAAATGAAA GCAAGTTATT ATGTAGGGAT TTTAAAGGAA ATTTTGTGAA	6480
20	AGTAAGTTTA TCATACACAC TTAATGTTGC GTATTGACGT TTAATGTTAG GTGTGTTCTT	6540
	TTATAGACGA TAAAAGCTGT GTGCATATTA AGCGAATGAT TTTCAAATTG ACGCTAATAT	6600
25	GCGAAAGTAG TATTTTTAAA ATGAACAACA ACGATGAAGA GGGGTTTATA GGATGAAAAT	6660
	TGCAATTGCT GGATCGGGTG CATTAGGTAG TGGCTTTGGT GCCAACTAT TTCAAGCAGG	6720
	ATATGATGTC ACACCTATTG ACGGATATAC ATCTCATGTT GAAGCGGTTA AGCAACATGG	6780
30	ATTAAATATA ACGATTAATG GAGAGGCATT CGAGTTAAAC ATTCCGATGT ATCATTTTAA	6840
	TGATCAACCG GACGAAAGCA TTTACGATGT TGTCTTTCTA TTTCCAAAGT CTATGCAATT	6900
	AAAAGAAGTG ATGGAAGATA TGAAGCCACA TATTGATAAT GAAACGATCG TCGTATGTAC	6960
35	GATGAATGGT CTGAAGCATG AAGAAGTCAT TGCGCAGTAT GTTGCTCAAT CACAAATTGT	7020
	CAGAGGTGTT ACGACTTGA CGGCAGGTCT TGAAAGCCCT GGACACAGTC ATTTACTTGG	7080
	TAGTGGACCA GTTGAAATAG GTGAACTAGT GGATGAAGGT AAAGAAAATG TTATAAAAGT	7140
40	TGCTGATTTA CTTAACGAAG CGGAATTGAA TGGTGTCAAT AGTAAAGATT TATACCAATC	7200
	GATTTGAAA AAGATTTGTG TTAATGGTAC GGCAAATGCA TTAAGCACAG TGTGAGTG	7260
45	TAATATGGCA TCGCTGAATG AAAGTAGTTA TGCGAAGTGT TTGATTTATA AATTAACGCA	7320
	AGAAATAGTG CATGTAGCGA CGATTGATAA TGTTCAATTA AATGTTGATG AAGTATTTGA	7380
	ATATTTAGTT GATTTAAATG AAaAAGTTGG TGCGCATTAT CCATCCATGT ATCAAGATTT	7440
50	AATTGTTAAT AATAGAAAAA CTGAAATTGA TTATATTAAT GCGCGAGTTG CAACATTAGG	7500
	TAAACAACGT CaTATTGAAG CGCCAGTCAA TCGCTTTATT ACTGATTTAA TTCATACTAA	7560

55

	CAATCACGTG ATATTACGGT CATTATTAAG ATTGAAATGT AATAAATAAA GAACAGCAGT	7680
	AAGGTACTTT CAAATTGAAA TGATCTTGGT GCTGTTTTTC TTGATTGATC TTCGTCATAA	7740
5	TTCAGATTTG TCATAGGCTA CGACATACTA TTAGTATTTA CTAGACAGTT TTTACGACGA	7800
	CACCTTGAAA AATTTTGAGG CAAATCATTT GGAAGTCTCA CGTGAATTTT GTAAACTCAT	7860
	CAAGCAAGTA ATTATATTAA AAAGACAAAT AGAGAAAAGG TGTTTATAAT GAGTAAAATT	7920
10	TTTGTAAGTGT GTGCAACGGG CCTTATTGGC ATTAAATTAG TTCAAAGACT AAAAGAAGAG	7980
	GGGCATGAGG TTGCTGGTTT TACTACATCT GAGAATGGTC AACAAAAGCT AGCTGCTGTT	8040
15	AATGTAAAAG CATATATTGG TGATATATTA AAAGCTGATA CTATTGATCA AGCGTTAGCA	8100
	GATTTTAAAC CAGAAATCAT TATCAATCAA ATTACGGATT TAAAAAATGT TGATATGGCA	8160
	GCAAATACGA AAGTACGTAT TGAAGGTTCT AAAAACCTAA TTGATGCGGC GAAAAAGCAT	8220
20	GACGTTAAGA AAGTAATTGC CCAAAGTATT GCCTTTATGT ATGAACCTGG CGAAGGATTA	8280
	GCAAATGAGG AAACCTCACT TGATTTTAAAC TCAACTGGCG ATAGAAAAGT AACGGTTGAT	8340
	GGTGTGGTTG GTTTAGAAGA AGAAACGGCT CGTATGGATG AATACGTTGT TTTACGTTTT	8400
25	GGCTGGTTAT ATGGCCCAGG TACTTGGTAC GGAAAAGATG GCATGATTTA TAATCAATTT	8460
	ATGGATGGTC AAGTGACACT TTCAGATGGC GTAACATCAT TTGTGCATCT TGATGATGCA	8520
	GTTGAAACAT CTATTCAAGC TATTCATTTT GAAAATGGTA TCTATAATGT AGCAGATGAT	8580
30	GCACCTGTTA AAGGTTCTGA ATTTGCAGAA TGGTATAAAG AACAACTTGG TGTTGAACCA	8640
	AATATTGATA TTCAACCTGC GCAACCATTT GAACGTGGCG TAAGCAATGA GAAGTTTAAA	8700
	GCGCAAGGTG GTACTCTGAT TTATCAAACCT TGGAAAGATG GCATGAATCC AATTAAATAA	8760
35	TAATTTATCC GTTTAATATA CAAAGAATAA AGACTTGGTC GAATCGTGGA TGATATATTA	8820
	TCAAACGCAC GGCTCGAACA AGTCTTTTTT ATTATGTCTT CGTTATCTTT GTATGAAGGA	8880
	ATAACAGAAT TACAATTAAT GTACTGAATA ATGCAATTAA TGTGTGATT AGTGCTAATT	8940
40	TAATTTCTAT TGGTAGCCAA GTCAGTACAA AAGACCAATT ATTGCTACCG AGAATGAGAT	9000
	ATGGTAATGC ATATAATATG AGCGCTAAAG CGATACATAT ACATAATGAT AACCAACTCA	9060
45	ATACAGCAAT CC	9072

(2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16826 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

	GTGGAACAGC TGTAACATA TCATTTCTTT CAACATTTAT TGGGAAAATG TTAGCTACAT	60
5	TTCTATATCC GATTAATAAT GTAGTACTTT CATATATnTC TGTAATGAA AGTGACAATA	120
	TAAAGAAGCA ATATTTGaaa ACTAATCTAA TTGCTATAGC TGCCCTATGT TTAGTCATGA	180
	TTATATGTTA TCCAATTACA ATAATTATTG TCTCTTTACT GTATAACATT GATTCAAGTT	240
10	TATATTCGAA GTTTATTATT TTAGGTAATA TAGGTGTTTT ATTCAATGCA GTGAGTATTA	300
	TGATCCAAAC TTAAATACA AAACACGCAT CAATAACATT ACAAGCGAAT TATATGACGC	360
	TTACACGAT TACATTTATA TTCATAACTA TTTTAATGAC AATTGCGTTT GGTCTAAATG	420
15	GATTCCTTTTG GACAACGCTG TTCAGCAACA TTATTAAGTA TGTGATTTTA AATATTATAG	480
	GTTTAAAGTC TAAATTCATT AATAAAAAGG ACGTCGATTA GATGAGTGAA AAAAAGATTT	540
	TGATTTTATG TCAGTATTTT TATCCGGAAT ATGTATCTTC TGCACGTTA CCAACTCAAT	600
20	TGGCGGAAGA TTAAATTGCG AATCACATTA ATGTCGATGT CATGTGTGGA TGGCCATATG	660
	AATATAGTAA TCATAAACAG GTTTCTAAAA CCGAGATGCA TCGTGGTATT CGCATTGAC	720
25	GTCTCAAGTA TTGAGGTTT AATAACAAAA GTAAGGTTGG AAGGATCATC AATTTCTTTA	780
	GTTTATTTTC AAAATTCGTG ATTAATATAC CTAAATGTT GAAATATGAT CAGATTCTTG	840
	TTTACTCTAA TCCACCAATC TTGCCATTAA TACCAGACGT TTTACACAGA CTGCTTAAGA	900
30	AAAAATATTC TTTTGTGGTG TATGATATAG CACCTGATAA TGCATTAAAG ACAGGTGCAA	960
	CTCGTCCAGG TAGCATGATT GATAAGCTGA TGCCTTACAT TAATAGACAT GTCTACAAGA	1020
	ATGCTGAAAA TGTCATTGTC CTTGGTACGG AAATGAAAAA CTACTTACTA AATCATCAAA	1080
35	TTTCTAAAAA TGCTGACAAT ATCCATGTGA TTCCTAACTG GTATGACATG CGTCAATTAC	1140
	AAGACAATCG TATCTATAAT GACACATTTA AAGCTTACCG TGAGCAATAC GACAAAATTT	1200
	TATTGTATAG CGGTAATATG GGGCAGTTAC AGGATATGGA GACACTTATC TCATTTTTAA	1260
40	AATTAAATAA GGATCAGTCT CAAACGTAA CAATACTTTG TGGTCATGGT AAGAAATTTG	1320
	CAGATGTCAA AACGGCAATA GaAGACCATC GTATTGAAAA TGTTAAAATG TTTGAGTTTT	1380
	TAACAGGTAC AGACTATGCT GACGTATTAA AAATTGCGGA TGTATGTATT GCATCGCTGA	1440
45	TTAAAGAAGG CGTCGGTTTA GCGGTGCCGA GCAAGAATTA TGGCTATCTT GCAGCTAAGA	1500
	AAGCGTTGGT ACTCATCATG GATAAGCAAT CTGATATCGT TCAACATGTT GAACAATATG	1560
50	ATGCGGGTAT CCAAATTGAT AATGGCGATG CACATGCCAT TTATAACTTC ATCAACACTC	1620
	ACTCGAGTAA GGAATTGCAC GAGATGGGTG AGCGCGACA TCAACTGTTT AAAGATAAAT	1680
55		

	AAGCGATTAT TCGATGTAGT GAGTTCAATA TATGGTTTAG TAGTTTTAAG TCCGATTCTG	1800
	TTAATTACAG CATTACTAAT TAAAAATGGAA TCACCTGGAC CAGCCATTTT CAAACAAAAA	1860
5	AGACCGACGA TTAATAATGA ATTGTTTAAT ATTTATAAGT TTAGATCAAT GAAAATAGAC	1920
	ACACCTAATG TTGCAACTGA TTTAATGGAT TCAACATCGT ATATAACAAA GACAGGGAAG	1980
	GTCATTGCGTA AGACCTCTAT TGATGAATTG CCACAATTAT TGAATGTTTT AAAAGGAGAA	2040
10	ATGTCAATTG TAGGTCCTAG ACCAGCGCTT TATAATCAAT ACGAATTAAAT CGAAAAACGT	2100
	ACAAAAGCGA ACGTGCATAC GATTAGACCA GGTGTGACAG GACTAGCTCA AGTGATGGGG	2160
	AGAGATGATA TCACTGATGA TCAAAAAGTA GCGTATGATC ATTATTACTT AACACATCAA	2220
15	TCTATGATGC TTGATATGTA TATCATATAT AAAACAATTA AAAATATCGT TACTTCAGAA	2280
	GGTGTGCATC ACTAATGAGA AAAAATATTT TAATTACAGG CGTACATGGA TATATCGGTA	2340
20	ATGCTTTAAA AGATAAGCTT ATTGAACAAG GACATCAAGT AGATCAAATT AATGTTAGGA	2400
	ATCAATTATG GAAGTCGACC TCGTTCAAAG ATTATGATGT TTTAATTCAT ACAGCAGCTT	2460
	TGGTTCACAA CAATTCACCT CAAGCAAGGC TATCTGATTA TATGCAAGTG AATATGTTGC	2520
25	TGACGAAACA ATTGGCACAA AAGGCTAAAG CTGAAGACGT TAAACAATTT ATTTTTATGA	2580
	GTACTATGGC AGTTTATGGA AAAGAAGGTC ATGTTGGTAA ATCAGATCAA GTTGATACAC	2640
	AAACACCAAT GAACCCCTACG ACCAACTATG GTATTTCCAA AAAGTTCGCT GAACAAGCAT	2700
30	TACAAGAATT GATTAGTGAT TCGTTTAAAG TAGCAATTGT GAGACCACCA ATGATTTATG	2760
	GTGCACATTG CCCAGGAAAT TTCCAACGGT TAATGCAATT GTCAAAGCGA TTGCCAATCA	2820
	TTCCCAATAT TAACAATCAG CGCAGTGCAT TATATATTAA ACATCTGACA GCATTTATTG	2880
35	ATCAATTAAT ATCATTAGAA GTGACAGGTG TGTACCATCC TCAAGATAGT TTTTACTTTG	2940
	ATACATCGTC AGTAATGTAT GAAATACGTC GCCAATCACA TCGTAAAACG GTATTGATCA	3000
	ACATGCCTTC AATGCTAAAT AAGTATTTTA ATAAGTTGTC GGTCTTTAGA AAATTATTCTG	3060
40	GCAATTTAAT ATACAGCAAT ACGTTATATG AAAATAATAA TGCACCTGAA ATTATTCCTG	3120
	GAAAAATGTC ACTTGTTATT GCGGACATCA TGGATGAAAC GACAACCAA GATAAGGCAT	3180
45	AAGTCATCTA TTAAATAAAA TCAACATACA AATCGTTTTA TTTGGAGGTT ATAGTATGAA	3240
	GTTAACAGTA GTTGGCTTAG GTTATATTGG TTTACCAACA TCAATTATGT TTGCAAAACA	3300
	TGGCGTCGAT GTGCTTGGTG TTGATATTAA TCAGCAAACG ATTGATAAGT TACAAAGTGG	3360
50	TCAAATTAGT ATTGAAGAAC CTGGATTACA AGAGGTTTAT GAAGAGGTAC TGTATCGGG	3420
	AAAATTGAAG GTATCTACAA CGCCAGATGC ATCTGATGTT TTTATCATTG CCGTTCCGAC	3480
55		

	TAGTATTTTA TCATTTTITAG AAAAAGGAAA TACCATTATT GTAGAGTCGA CAATTGCGCC	3600
	TAAAACGATG GATGATTTTG TAAAACCACT CATTGAAAAT TTAGGGTTTA CAATAGGTGA	3660
5	AGATATTTAT TTAGTGCAAT GTCCAGAACG TGTACTGCCA GGAAAAATTT TAGAAGAATT	3720
	AGTTCATAAC AATCGTATCA TTGGCGGTGT GACTGAAGCT TGTATTGAAG CGGGTAAACG	3780
	TGTCTATCGC ACATTCGTTT AGGGAGAAAT GATTGAAACA GATGCACGTA CTGCTGAAAT	3840
10	GAGTAAGCTA ATGGAAAACA CATATAGAGA CGTGAACATT GCTTTAGCTA ATGAATTAAC	3900
	AAAAATTTGC AATAACTTAA ATATTAATGT ATTAGATGTG ATTGAAATGG CAAACAAACA	3960
	TCCGCGTGTT AACATCCATC AGCCTGGTCC AGGTGTAGGC GGTCATTGTT TAGCTGTTGA	4020
15	TCCGTACTTT ATTATTGCTA AAGACCCTGA AAATGCAAAG TTAATTCAA CTGGACGTGA	4080
	AATTAATAAT TCAATGCCGG CCTATGTTGT TGATACAACG AAGCAAATCA TCAAAGTGTT	4140
20	GAGCGGGAAT AAAGTCACAG TATTTGGTTT AACTTATAAA GGTGATGTTG ATGATATAAG	4200
	AGAATCACCA GCATTTGATA TTTATGAGCT ATTAATCAA GAACCAGACA TAGAAGTATG	4260
	TGCTTATGAT CCACATGTTG AATTAGATTT TGTGGAACAT GATATGTCAC ATGCTGTCAA	4320
25	AGACGCATCG CTAGTATTGA TTTAAGTGA CCACTCAGAA TTAAAAATT TATCGGACAG	4380
	TCATTTTGAT AAAATGAAGC ATAAAGTGAT TTTTGATACA AAAAATGTTG TGAAATCATC	4440
	ATTTGAAGAT GTATCGTATT ATAATTATGG CAATATATTT AATTTTATCG ACAAATAAAA	4500
30	TGTGTCAAAC TAGGCATAC ATGATTAAGG AAAGATAAGC TGTGATGTGT TTGAACCTCA	4560
	GAGAGGATAA TGTTATGAAA AAAATTATGG TTATTTTCGG TACGAGACCC GAAGCAATAA	4620
	AAATGGCACC ATTAGTAAAA GAAATTGATC ATAATGGGAA CTTTGAAGCG AACATTGTGA	4680
35	TTACAGCACA ACATAGAGAT ATGTTAGATA GTGTGTTAAG TATATTTGAT ATTCAAGCTG	4740
	ATCATGATTT AAATATTATG CAAGATCAAC AAACATTAGC AGGCCTTACG GCGAATGCAC	4800
	TTGCTAAACT TGATAGCATC ATTAATGAGG AACAACCGGA TATGATTTTA GTACATGGTG	4860
40	ATACTACAAC GACTTTTGTA GGAAGTTTGG CAGCATTTTA TCATCAAATT CCGGTCGGAC	4920
	ATGTAGAAGC TGGACTTCGA ACACATCAGA AATACTCACC ATTCCTGAA GAGTTAAATC	4980
45	GAGTCATGGT AAGTAATATT GCTGAATTGA ATTTTGCGCC AACAGTAATT GCAGCTAAAA	5040
	ATTTACTTTT TGAACAACAA GACAAAGAGC GTATCTTTAT TACTGGAAAT ACAGTTATTG	5100
	ACGCATTGTC AACAACAGTT CAAAATGATT TTGTTTCAAC GATTATTAAT AAACATAAAG	5160
50	GCAAGAAAGT TGTTTTACTA ACAGCGCATC GTCGTAAAA TATTGGGGAA CCGATGCATC	5220
	AGATTTTTAA AGCAGTAAGA GATTTGGCAG ATGAATATAA AGATGTTGTC TTCATTTATC	5280

55

	GGATTGAATT AATTGAGCCA TTAGATGCGA TTGAGTTCCA TAATTTTACA AATCAATCGT	5400
	ACCTCGTGCT GACAGATTCT GGTGGTATTC AAGAGGAGGC TCCTACATTT GGAAAAACCTG	5460
5	TGTTGGTATT AAGGAATCAT ACAGAGCGTC CCCAAGGCGT TGAGGCGGGA ACATCGAGAG	5520
	TAATTGGCAC AGATTATGAC AATATTGTTC GAAATGTGAA ACAATTGATT GAGGATGATG	5580
	AAGCGTATCA ACGTATGAGT CAAGCGAATA ATCCATATGG TGATGGACAA GCATCACGAC	5640
10	GTATTTGTGA AGCAATAGAA TATTATTTTG GATTGCGCAC AGACAAGCCG GATGAATTCG	5700
	TACCTTTACG TCACAAATAA TAAAAAACCC CTAATCATGA AGTTGGTTTA GACAACCAGC	5760
15	GGTGACTAGG GGTTTTTAAT ATATTTATTT TTGATAGTGG TAGCCAATAT CATATTTGAA	5820
	TACTTTATTT GATAATATTG GACTTTGCTG TCCATCGTCA TCACTTTTAA AACGTACATT	5880
	TTTATGAGCT TCTTTAAATA CATCGGAATT CAACCAATTA TTAAAGCTAT CTTCAGATTC	5940
20	CCAAATAGTT AAGATTTTAA CTTCGTCTGT ATCCTCGGTA TTTAATGTTT TAGTGACAAA	6000
	CATTTGTTGG AAGCCTTCAA TAGTTTCAAT ACCTTGCTA TTGTAAAAAC GTTCAATCGT	6060
	TTCTTCCGCA CTGCCTTTTT GTAATTGTAA TCTATTTTCT GCCATAAACA TGGGCAATCA	6120
25	CTCCTCTATT TTATGATTG ATTTGGGTAA TGTTTTTACA AATGTAAAGA GTACAGCGGT	6180
	TTGTATGATA ACCATTATGA TTAATCCTAC ACGGACTGCA AGAACATCCA CCATATAAAT	6240
	TGAAAAACCT ATTACAATGT ATAAGCTAAT TAAAATTTTA ATTTTCTGTT GTAGCGTGTA	6300
30	GCCTCGATGT AAATAAAAGT TTTCTACATA TTCTTTATAA ATTTTTTGAT TAATAAGCCA	6360
	ATTGTAAAAG CGATCTGAAC TTCGAGCAAA GCAAAAAACT GCTACGAGTA AAAAAGGGGT	6420
	CGTTGGCAGT AAAGGTAATA CGGCACCTGC AATACCAAGC GCTGTAAATA TTAAGCCAAT	6480
35	GACGATTAAA ATAAGTCGCA TTGAAAAAAC TCCATTCTAG TACTAATGCG CATGTAATAT	6540
	TGTTTTAGTA ATATAACTCA TGCTAAATAT AATGTGTATG ATAAGTGCAA TGA CTGAGTA	6600
	AAATGAAACG ATGTTGAATT ATCCTTGTC AATTAAACGCA TTTTAAGCGC GACTTTCATA	6660
40	ACAACCAAAC TATTTAATGA GAATTATTCT CAAGTATTAT AGTTATATTA TGTGTTTTAT	6720
	TTTTGAAAAG TGCAATATGT TTTGAAAAT AAGATTATTT TTATGTGCAA AAACGACGCA	6780
45	AAAGTTTTAA AAATGAGACT TCTGTGAGCT GATTATTTTA TAAAATGTAA ACGCTTACTA	6840
	TATAATGTGA ATCATATCGT TTAAAAGCAT TATTAAATAT GATGCTAAGA GATTTATATT	6900
	ATAGCCAATA AACAAAGGAG AGATAATATG GCAGTAAACG TTCGAGATTA TATTGCAGAG	6960
50	AATTATGGTT TATTTATCAA TGGGGAATTT GTTAAAGGTA GCAGTGACGA AACAAATCGAA	7020
	GTGACTAATC CAGCAACTGG AGAAACACTA TCACATATTA CAAGAGCAAA AGATAAAGAT	7080

55



	TCAGAACGTG CACAAATGTT GCGTGATATT GGTGATAAAT TAATGGCACA AAAAGATAAA	7200
	ATTGCAATGA TTGAAACATT AAATAATGGT AAACCGATTG GTGAGACAAC AGCAATTGAT	7260
5	ATTCCATTG CTGCAAGACA TTTCCATTAT TTCGCAAGTG TTATTGAAAC AGAAGAAGGT	7320
	ACAGTGAATG ATATCGATAA AGACACAATG AGTATCGTAC GACATGAGCC GATTGGCGTC	7380
	GTAGGTGCTG TTGTTGCTTG GAACTTCCCA ATGCTATTAG CTGCATGGAA GATTGCGCCA	7440
10	gCCATTGCTG CAGGTAATAC AATTGTGATT CAACCTTCGT CTTCAACACC ATTAAGTTTA	7500
	TTGGAAGTTG CTAAATTTT CCAAGAGGTA TTACCTAAAG GTGTTGTCAA TATACTAACG	7560
15	GGTAAAGGTT CAGAATCAGG TAATGCAATT TTCAATCATG ATGGTGTAGA TAAATTATCA	7620
	TTTACGGGCT CAACTGATGT AGGTTATCAA GTTGCCGAAG CTGCAGCAA ACATCTAGTA	7680
	CCCCTACAT TAGAGCTTGG TGGTAAAGC GCCAATATCA TATTAGATGA TGCTAATTTA	7740
20	GACCTTGCGT TTGAAGGTAT TCAGTTAGGT ATTTTATTC ACCAAGGTGA AGTATGTAGT	7800
	GCAGGTTCTC GATTATTAGT TCATGAAAAA ATTTATGATC AATTGGTGCC ACGTTTACAA	7860
	GAGGCATTTT CAAATATTAA AGTTGGAAAT CCACAAGATG AAGCTACACA AATGGGTAGT	7920
25	CAAACCTGGTA AGGATCAATT AGATAAAATT CAATCATATA TTGATGCAGC AAAAGAATCA	7980
	GATGCACAAA TTTTAGCAGG CGGTCATCGC TTAAGTAAA ATGGATTAGA TAAAGGGTTC	8040
	TTCTTTGAGC CGACATTAAT TGctGTGCCA GACAATCATC ACAAATTAGC ACAAGAAGAA	8100
30	ATATTTGGAC CAGTGTTAAC AGTGATTAAA GTGAAGGACG ATCAAGAAGC AATTGATATA	8160
	GCTAATGATT CTGAGTATGG TTTAGCAGGC GGTGTATTTT CTCAAAATAT CACACGTGCA	8220
	TTAAATATTG CTAAAGCTGT ACGTACAGGA CGTATTTGGA TTAACACTTA CAACCAAGTA	8280
35	CCAGAAGGCG CACCATTGCG TGGTTATAAA AAATCAGGTA TCGGTCGAGA AACTTATAAA	8340
	GGTGGGTTAA GTAACATCA ACAAGTTAAA AATATTTATA TTGATACAAG CAATGCTTTA	8400
	AAAGGTTTGT ACTAGAATAA ATATCGTTTC TGAAGCGTGT TTGTAGGTCA GTCTAGCGGT	8460
40	AAGTCTTAAC ATTTAACGGC GTTGTTTAGA TTTTAAGCAA AACAAAATAT ATAGGAACAC	8520
	GTATCATGAT ATTAGGATAT AATGACTAAA ATAATAGCAG TAGGATGGTT TTTAATTGCA	8580
45	AATCATCTTA CTGCTGTTTT TAATTATGCT AATTTGCGAT GCGGCTATTA TAAGGACAGA	8640
	GTTGTTTATT AATTATGGTG ATTTAGAAAT ATGAAGTTCA ATATGCAAAG TCATCGTTTG	8700
	TTTTAATATG CGGAACAATC ATTAAAGTTA TTGCGATTTT TTGAACTTAA TGAAACTAAA	8760
50	CAATAAATTT GAGATACTTT TTTGTCAATTT TTATGTAAGT AACACAATAA TCTCGTACAT	8820
	TATTAAAAAT TTCTATATGA TAGGAATAAA GCAAAGCGCG AGTGTGCTGT AAAAGTTTTT	8880
55		

	GATGATGTAT AAATCATGGT TAATTACGGA AGCATTAAATA TTAACCTGAG AAGCTATAAA	9000
	GAATTATTTT TAAAAGCGAC AATATTAAAT ACGACGCATT TATTTAGGAG TGGCAAACGT	9060
5	ATGAATGGGA AAAAGGCGAA TACGATAAAC AGATACAAAT ATTTTCATCA TGTCAATCAT	9120
	CAAAAAATTC AACAAAGTTC TAAAAAGACG CTGTGGGCAT CACTAATCAT CACATTGTTA	9180
	TTTACAGTGA TTGAATTTGT CGGAGGTTTA GTATCTAATC CATTGGCATT ACTGTCAGAT	9240
10	TCATTTTATA TGCTTAGTGA TGTATTAGCA CTTGGTTTAT CTATGTTGGC CATTTATTTT	9300
	GCAAGTAAAA AGCCGACTGC ACGATACACA TTTGGATATT TAAGATTTGA GATATTAGCT	9360
	GCATTTTAA ATGGTTTAGC ATTAATTGTA ATTTCAATCT GGATTTTATA TGAAGCTATT	9420
15	GTACGTATTA TTTATCCGCA ACCAATTGAA AGTGGCATT TGTATTATGAT TGCTAGTATT	9480
	GGTTTACTCG TCAATATTAT TTTGACTGTT ATCCTTGTA GGTCTTTAAA ACAAGAAGAC	9540
	AATATCAATA TTCAAAGTGC ATTATGGCAT TTCATGGGAG ACTTATTGAA CTCTATTGGT	9600
20	GTCATCGTTG CAGTTGTATT GATTTACTTT ACAGGATGGC GCATCATCGA CCCAATCATT	9660
	AGTATTGTAA TTCACTCAT CATTTTACGT GGTGGTTATA AAATTACGCG TAATGCgTGG	9720
25	CTAATTTTAA TGGAAAGTGT GCCTCAACAT TTGGATACTG ATCAAATTAT GGCAGATATT	9780
	AAAAACATAG ATGGCATATT AGATGTACAT GAATTTTCATT TGTGGAGTAT TACAACAGAG	9840
	CATTATTCAT TAAGTGCCCA TGTGTGTTA GATAAAAAAT ATGAGGGTGA TGATTATCAA	9900
30	GCGATTGATC AAGTATCATC ATTGTTGAAA GAAAAATATG GCATTGCACA TTCAACGTTG	9960
	CAAATTGAAA ACTTGCAATT GAATCCATTA GATGAGCCAT ACTTCGACAA ATTAACATAA	10020
	ATAAACATT GTAGCGCCTA AAACATTAAT CTATGTCATA GCGCACGTT TCGTTTTATA	10080
35	CTTATGTTGC ATCATTTAAA TGATTTTCGT CAATTTCTTT GATGCTATCT ACATCTAACA	10140
	CGACATCTTT AGGTTTCAA ATATGAATAT GTTTTTCATC ATTTGTATGT AAAATGCGTT	10200
	CTATGATGTA CCTTTGACCG GCCATTGTTT CTACAGCAAT CTTTTTGTTT CTAGCTAAAC	10260
40	TTGCTACGAC AGATTCTTTA TCCATAATGA TAGCCCCCTA TATATATGTT TATTTACTTA	10320
	TACCCTAACA TGATTTTTAT ACTCTTGAA AATATATTTT ACAGAATTTT ATCTAAATAT	10380
	TTAAAAAAT ATCTTAATAT CCTTGTAATC CGATAAGAAT TATAGTAATA TTTTTTCAAC	10440
45	CATcGTTATA GGAGGTCTTA TTAATGACAT TATTTTTATT AGAAGCTAAC AATCTTGATT	10500
	TTGCATCAAC GAAAGAAGAA CTAGAAGCAA AGGCAGCATC ACTATCTACG AAGACAATTC	10560
	CAACATTAAT TGAAGTACAA GCTACTGAAA ATTTAACTCA TGGTTATTTT ATTGTGGAAG	10620
50	CAAATGACGA aGCAGAAGCT AAACAATTTT TAACAGAAGC AGATATTAGT ATTCAATTAG	10680

55

	TTGATTACCT TGTAACCTGG AACATTCCGG AAGGCATTAC GATGGATCAA TATTTAGCAC	10800
	GTAAAAAGAA AAATTCTGTT CATTATGAAG AAGTGCCAGA AGTTGAATTT AAACGCACAT	10860
5	ATGTATGTGA AGATATGTCT AAATGTATTT GTTTATACAA CGCACCTGAT GAAGAAGCGG	10920
	TACGTCGCGC GCGCAAAGCA GTTGATACAC CGATTGATGG CATCGAAAAA CTTTAATAAG	10980
	ACAACAAGTT GATGAGATAT ATGTATATAG GTTTGGCATG GATTTTCGATT GCAGTTAATT	11040
10	AGAATAGCTC AATGCTATAA ATGTAAGTAG TTGATATGAA GAACTAATG AACTAAATGC	11100
	AAGTATTGTC TAAAAACAAT ATTTTATTGA AATTTAGTAG AGCTGAAATT AATATAACGT	11160
	CGTTAATTGA ATAACGCTTA TGTTATAAGA GCACTCATAC CAAACCATAA TCATCTATAG	11220
15	ATATAACAAT TCACGATATA AGGGCTGTGT TTGGCATAGC CCTTTAGATA TACACTTAAT	11280
	TCCTATTAAA ATAGTAGGGA TTAAGGGG GCTTGTCATG ATTAATAATC AACAAATTACA	11340
	ACATCACTTT GGATCACATA AAGTAATTCA TAACTTTAAT TTGGACATTA GCAAGGGAGA	11400
20	AATAGTCACT TTCATAGGGA AAAGTGGTTG CGGAAAGTCT ACTTTACTCA ATATTATCGG	11460
	TGGATTTATT CATCCATCGT CTGGTCGTGT CATTATTGAT AACGAAATTA AACAAACAGCC	11520
25	ATCTCCAGAT TGTTTAATGC TATTTCAACA TCATAATTTG CTGCCATGGA AAACGATTAA	11580
	TGACAACATT AGGATTGGAT TACAACAGAA AATTAGTGAT GAAGAGATTA ACGCACAGCT	11640
	TAAATTAGTT GATTTAGAAG ACAGGGGAAA GCATTTTCCC GAGCAACTGT CCGGGGGTAT	11700
30	GAAACAACGT GTGGCACTAT GTCGAGCGCA TGTGCATAAG CCTAACGTTA TATTGATGGA	11760
	TGAGCCATTA GGTGCATTAG ATGCATTTAC ACGTTATAAA CTTCAGGATC AACTAGTGCA	11820
	aCTAAACAT AAAACGCAAT CAACTATTAT TTTAGTGACG CATGACATTG ATGAAGCTAT	11880
35	TTATCTTTCC GACCGCATTG TTCTGTTAGG TGAAGGGTGC AATATTATTT CTCAATATGA	11940
	AATTACAGCA TCACATCCAC GCAGTCGTAA TGATAGCCAC CTACTTAAGA TTCGTAATGA	12000
	AATTATGGAA ACATTTGCAT TGAATCATCA TCAAGTTGAA CCTGAATATT ATTTATAAGG	12060
40	AGTGAGTGAC GATGAAAAGG TTAAGCATAA TCGTCATCAT TGAATCTTT ATAATTACAG	12120
	GATGTGATTG GCAAAGGACG TCTAAAGAAC GGTCTAAAAA TGCCCAAAAT CAGCAAGTGA	12180
	TTAAAATTGG ATATTTGCCG ATTACACATT CAGCTAATTT GATGATGACT AAAAAATTAT	12240
45	TATCACAATA CAATCATCCG AAATATAAAC TAGAATTAGT TAAATTCAAT AATTGGCCAG	12300
	ATTTAATGGA CGCATTAAAC AGTGGTCGTA TTGATGGTGC ATCAACTTTA ATAGAGCTAG	12360
	CGATGAAATC AAAACAGAAG GGCTCAAATA TAAAGGCTGT GGCAATTGGGC CATCATGAAG	12420
50	GCAATGTCAT TATGGGACAA AAAGGTATGC ACTTAAATGA ATTTAATAAT AATGGCGATG	12480
55		

	GTAAACAATT AAAGATTAAA CCGGGGCATT TTAGCTATCA TGAAATGTCG CCAGCAGAAA	12600
	TGCCAGCCGC ATTGAGTGAA CACAGAATTA CAGGGTATTC TGTAGCCGAA CCATTCCGGT	12660
5	CACTGGGTGA AAAGTTAGGC AAAGGTAAGA CTTTGAAACA TGGTGATGAC GTTATACCTG	12720
	ATGCGTATTG CTGTGTGCTA GTACTGAGAG GGGAAATTGCT TGATCAACAC AAGGATGTAG	12780
	CGCAAGCATT TGTACAAGAT TATAAAAAGT CTGGCTTTAA AATGAATGAT CGCAAGCAAA	12840
10	GTGTAGACAT TATGACGCAT CATTTTAAAC AAAGTCGTGA CGTTTTAACA CAGTCAGCGG	12900
	CATGGACATC CTATGGTGAT TTAACAATTA AGCCATCCGG CTATCAAGAA ATTACGACAT	12960
	TGGTAAAACA ACATCATTTG TTTAATCCAC CTGCATATGA TGACTTTGTT GAACCGTCAT	13020
15	TGTATAAGGA GGCATCGCGT TCATGACACG TCCACAAAT AACAAATTTA TATTACCTAT	13080
	TATCACATTT ATTATTTTCT TAGGCATTG GGAATGGTC ATTATTATTG GGCATTACCA	13140
	ACCTGTATTG TTACCGGGTC CTGCTCTTGT AGGAAAAAGT ATATGGTCTT TCATTGTTAC	13200
20	TGGAGAAATT TTCCAACATT TAGCAATTAG TTTATGGAGA TTTGTAGCGG GCTTTGTTGT	13260
	CGCATTGTTG GTTGCTATTC CATTGGGCTT CTTGCTTGGA AGGAATCGTT GGCTATACAA	13320
25	CGCTATCGAA CCGCTATTTT AATTGATTAG GCCGATATCT CCGATAGCAT GGGCACCATT	13380
	TGTTGTCTA TGGTTTGGTA TTGGTAGTTT GCCAGCGATT GCGATTATTT TTATCGCTGC	13440
	TTTTTTCCCA ATTGTGTTCA ATACTATTAA AGGCGTTAGA GACATTGAAC CTCAATATTT	13500
30	AAAAATAGCA GCAAATTTAA ATTTAACTGG GTGGTCATTG TATCGCAATA TATTATTTCC	13560
	CGGGGCATTT AAACAAATCA TGGCTGGGAT ACATATGGCG GTAGGAACAA GTTGGATATT	13620
	TTTAGTTTCT GGTGAAATGA TTGGTGCACA ATCGGGATTA GGTTTTTTAA TCGTTGATGC	13680
35	ACGAAATATG TTGAACCTAG AAGATGTTTT AGCAGCAATA TTCTTTATCG GATTATTTGG	13740
	TTTTATTATT GATCGATTCA TTAGTTATAT TGAGCAGTTT ATACTTAGAA GATTGGTGA	13800
	ATAAGGAGAG ATGATGATGA CTTTAGAAAC GCTTATCAAA GAACAATTAG ATCCTCATTT	13860
40	AGTAGAAGTT GATGAAGGGA CGTATTATCC GAGAACATTT ATTCAGCAAT TATTTGTAGA	13920
	TGGTTATTTT GGTGAGGCGG CATTGAGAAA AAATGCTGAA GTAATCGAAG CTGTATCGCA	13980
	GTCTTGTTTG ACAACAGGAT TTTGTTTATG GTGCCAATTA GCTTTTTCAA CGTATTTAGA	14040
45	AAATGCCACG CAGCCACATT TAAATAATGA CTTACAACAG CAATTGTTAT CTGGAGAAAT	14100
	ATTAGGTGCT ACCGGATTGT CTAATCCGAT GAAGTCATTT AATGATTAG AAAAGTTGAA	14160
	CCTTGAACAC ACTTATGTTG ATGGACAATT GGTGTGAGT GGACGTATGC CAGCTGTAAG	14220
50	TAATATTCAA GAAGACCATT ATTTTGGTGC GATTTCGAAA CATGAATCAT CAGATGAATT	14280

55

	TTTAGGAGTC AACGGGTCAG CAACGTATCA AATCACATTG AATCAAGTCG TAGTGCCACA	14400
	ATCACAAATT ATCAGCATG ATGCGAAGCA GTTTGCGGCA ACTATTCGCC CGCAATTTAT	14460
5	TGCTTACCAA ATTCCAATAG GATTAGGCTC AATTAAAAGT TCTTTAGAGT TAATTGATGC	14520
	ATTTTCAAAT GTGCAAAACG GAATAAATCA ATATTTAGAG TATGATGTTG AAGCTTTTAA	14580
	AAAACGTTAT CGTCAACTTA GAGAGGAATA TTATGCAATA TTAGATGACG GTAACCTAAC	14640
10	TTACATTTA AATGAATTAA TATCATTGAA GAAGGACATC GGCTATTTAT TGTTAGATGT	14700
	AAATCAAGCT TCTGTTGTCA ATGGTGGTTC TAGAGCGTAC ACACCATATT CGCCACAAGT	14760
	TCGCAAGTTA AAAGAAGGAT TCTTCTTCGC AGCATTGACA CCGACATTAA GACATTTAGG	14820
15	TAAACTTGAA GCAGAGTTGA AGGGGTAAAGT GTGATAAGCT GATTTTTTGT TTAGATGCGT	14880
	TTGTTGAAAC ATTTTTTAAA ATAATATAAA TCTTAGTTTA TAAACATTTT CTGTTAATTT	14940
20	GTTATATCCT TTAACTAGG AAAATATACA TTTCGTAATA ATAATAATCG TTATCATTGA	15000
	AAAAGTGTTA ATAAGGTGTA TAATGAAAAT GTGAACAATT AATGAACCTC TTATTTTAAA	15060
	GAAGGTGAAT ACTATAGATA CGCATACTAA AGAACAACAA TTCTCGAATC TAGTAAGATC	15120
25	TTATCGTAAA GAATACGTGG GTAAAGGACC CAATAGTATT CGAGTGTCTG TTAAGATAA	15180
	TTGGGCGATT GCACATATGA CAGGTGTTTT GAGTAAAGTT GAGAGTTTTT ACCTAAACGA	15240
	CAAACGCAAT GAATCGATGC TCCATTATAC ACGCACAGAG AAGATTAAAC AGATGTATAA	15300
30	AGAAATAGAT GTAAATGAGA TGGAAAGTCT TGTAGGCGCT AAGTTTGTAA AATTATTTAC	15360
	AGATATTGAT TTGAATGATG ATGAAGTCAT TTCAATATTT GTTTTCGATA AGTCAATAGA	15420
	ATAAGTGTG CTGGTGTAAG GTACACGGTG CTGTTTGCTA ACTTCGCTTT GAATTAAACA	15480
35	ATAATTCAAG GGGGTGGTAT GTCAAACGGT GCCGTTTTTT TGTCATATTT TTAACAACAG	15540
	CAACATGCAA CACGTACTTT AAGGAAGTCA AAATTTATCA TTTAGGAGAG ATGGATATGA	15600
	AAATCGTAGC ATTATTTCCA GAAGCAGTAG AAGGTCAAGA AAATCAATTA CTTAATACTA	15660
40	AAAAAGCATT AGGATTAAAA ACATTTTTAG AGGAAAGAGG ACATGAGTTC ATTATATTAG	15720
	CAGATAATGG TGAAGACTTA GATAAACATT TACCAGATAT GGATGTGATT ATTAGTGCGC	15780
45	CATTTTATCC TGCATATATG ACTCGTGAAC GTATTGAAAA AGCACCGAAC TTGAAATTAG	15840
	CAATTACAGC AGGTGTAGGA TCTGACCATG TAGATTTAGC GGCAGCAAGT GAACACAATA	15900
	TTGGTGTCTG TGAAGTTACA GGAAGTAATA CAGTTAGTGT GGCAGAACAT GCGGTTATGG	15960
50	ATTTATTAAT ACTTCTTAGA AACTATGAAG AAGGTCATCG TCAATCAGTA GAAGGTGAAT	16020
	GGAAGTTGTC TCAAGTAGGT AATCATGCGC ATGAATTACA ACACAAAACA ATTGGTATTT	16080
55		

TACAACACTA TGATCCAATC AATCAACAAG ACCATAAATT GTCTAAATTT GTAAGCTTTG 16200  
 ATGAACCTTG GTTCAACAAGT GATGCGATTA CAATTCATGC ACCATTAAACA CCAGAAACTG 16260  
 5 ATAACCTATT TGATAAAGAT GTTTTAAGTC GTATGAAAAA ACACAGTTAT TTAGTGAATA 16320  
 CTGCACGTGG TAAAAATTGTA AATCGCGATG CGTTAGTTGA AGCGTTAgCA TCCGAGCATT 16380  
 TACAAGGATA TGCTGGTGAT GTTTGGTATC CaCAACcTG ACCTGCTGAT CATCCATGGA 16440  
 10 GAACAATGCC TAGAAATGCT ATGACGGTTC ACTATTCAGG TATGACTTTA GAAGCACAAA 16500  
 AACGTATTGA AGATGGAGTT AAAGATAATTT TAGAGCGTTT CTTCAATCAT GAACCTTTCC 16560  
 AAGATAAAGA TATTATTGTT GCAAGTGGTC GTATTGCTAG TAAAAGTTAT ACAGCTAAAT 16620  
 15 AGAATAAGGA TGCTGGGCTA GCGATTAAAG CTTTCAATTT TATATAAATG AATCATATAA 16680  
 GCACTACTGC TGTTGTAAAG ATGGCAGTAG TTTTTTTATG ATTACATCTA AGTATAGTCA 16740  
 CGGTATGTT AGGACAATGA TTTAACATTT ACGCACATAT GTGTTCACTT ACGCAATTAT 16800  
 20 TGAnAAATnT CATTCAATGTG GnAATC 16826

(2) INFORMATION FOR SEQ ID NO: 47:

25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 4012 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

TTCAATGAGA GTAGTGGGCT GATGTTTAGC GATATCGCGT AAGATTAAAC ATTGGCCATA 60  
 35 ATATATATTG TGTTTTTCTA AAATCGGCTC GGCTAATTTT AAATAGGGGC GATATATTGT 120  
 TATAAACTA TTGAAAAATT CTTGTGATAG CATAGTGACA TCTCCTAAGA CAAAATAGTT 180  
 AGCTTAGCTA mCCTTTTTTAC AACAAATAGTA ATTATAAAC GGGAGCAATT AGAAATCAAT 240  
 40 ATATAATTAT TAAGAGCAAA AATAATTATA CTTTGTTAAA ATAAGCGTAA TTACATGTAA 300  
 ATAGGGGGAT ACTAATGATA TTGAAATTTG aTCACATCAT TCATTATATA GATCAGTTAG 360  
 ATCGGTTTAG TTTTCAGGA GATGTTATAA AATTACATTC AGGTGGGTAT CATCATAAAT 420  
 45 ATGGAACATT CAATAAATTA GGTATATCA ATGAAAATTA TATTGAGCTA CTAGATGTAG 480  
 AAAATAATGA AAAGTTGAAA AAGATGGCAA AAACGATAGA mGGCGGAGTC GCTTTTGCTA 540  
 CTCAAATTGT TCAAGAGAAG TATGAGCAAG GCTTTAAAAA TATTTGTTTG CGTACAAATG 600  
 50 ATATAGAGGC AGTTAAAAAT AAACACAAA GTGAGCAGGT TGAAGTAGTA GGGCCGATTC 660

55

	ATCAGGATGA TGATGAAATT AAGCCACCAT TTTTATTCA ATGGGAAGAA AGTGATTCCA	780
	TGCGTACTAA AAAATTGCAA AAATATTTTC AAAAACAATT TTCAATTGAA ACTGTTATTG	840
5	TGAAAAGTAA AAACCGATCA CAAACAGTAT CGAATTGGTT GAAATGGTTT GATATGGACA	900
	TTGTAGAAGA GAATGACCAT TACACAGATT TGATTTTAAA AAATGATGAT ATTTATTTTA	960
	GAATTGAAGA TGGTAAAGTT TCAAAATATC ATTCGGTTAT CATAAAAGAC GCACAAGCAA	1020
10	CTTCACCATA TTCAATTTTT ATCAGAGGTG CTATTTATCG CTTTGAACCA TTAGTATAAA	1080
	TATACGTAAG TGCTATGAGC GAGAATGCCC ATATGAATAA TGACAAGCAC AATGGAAAGA	1140
	ATCGTTAATA TATTATTTAA TCGTGATGAC TTAATTAAAA TGAAAAAGAT TGATAATATA	1200
15	AATGTGAAAA AGATAAGTAT AACCCGTAAT CTAAAGTAAT TCACGGTGAG AGGTTGACTC	1260
	AATGTCATAA TGATTGCAAC GATGTTTATA ATTATAAATA GACTTAAAAT AATTGTTCTC	1320
	ATATCAAACA CCTCATTGTT AGATTATTGA CATTATAACA GGGGTAATTG TATATGAACA	1380
20	TTAATGTGGT TGCTTGAGGA AAAATTTATT CATTGAAGTC AAGTTGGTTC ATTTTAGAAA	1440
	TGAATATCGT GTTAGATGAT GAAAGTATAT TGAAGTATAG GTAAC TAGTT GAAAAGTATT	1500
25	AATTGTACGA TAACATTAAA TTTAACACGA AACATAGATA TAAAATGATT CACAATTAAA	1560
	ATGGGTAAAT TTGAACTGTC TAAACTATTA ATTGGAGCAT GGACATTTCA AAAATAAGAG	1620
	TTCAAATCTT ACACAAGCTC TGAATCGACA CTATAAGATA CAAACTGTAT AATTAAAGGT	1680
30	ATTGTTAAAT AGAAGGAGAT ATCATAAATC ATGGAAAAGA TGCATATCAC TAATCAGGAA	1740
	CATGACGCAT TTGTTAAATC CCACCCAAAT GGAGATTTAT TACAATTAAAC GAAATGGGCA	1800
	GAAACAAAGA AATTAAGTGG ATGGTACGCG CGAAGAATCG CTGTAGGTCG TGACGGTGAA	1860
35	GTTCAGGGTG TTGCGCAGTT ACTTTTAAAA AAAGTACCTA AATTACCTTA TACGCTATGT	1920
	TATATTTTCGC GTGGTTTTGT TGTTGATTAT AGTAATAAAG AAGCGTTAAA TGCATTGTTA	1980
	GACAGTGCAA AAGAAATTGC TAAAGCTGAG AAAGCGTATG CAATTAAAAT CGATCCTGAT	2040
40	GTTGAAGTTG ATAAAGGTAC AGATGCTTTG CAAAATTTGA AAGCGCTTGG TTTTAAACAT	2100
	AAAGGATTTA AAGAAGGTTT ATCAAAAGAC TACATCCAAC CACGTATGAC TATGATTACA	2160
	CCAATTGATA AAAATGATGA TGAGTTATTA AATAGTTTTG AACGCCGAAA TCGTTCAAAA	2220
45	GTGCGCTTGG CTTTAAAGCG AGGTACGACA GTAGAACGAT CTGATAGAGA AGGTTTAAAA	2280
	ACATTTGCTG AGTTAATGAA AATCACTGGG GAACCGCATG GCTTCTTAAC GCGTGATATT	2340
	AGTTACTTTG AAAATATTTA TGATGCGTTG CATGAAGATG GAGATGCTGA ACTATTTTAA	2400
50	GTAAAGTTGG ATCCAAAAGA AAATATAGCG AAAGTAAATC AAGAATTGAA TGAACCTCAT	2460

55

CAAAATATGA TTAATGATGC GCAAAATAAA ATTGCTAAAA ATGAAGATTT AAAACGAGAC 2580  
 CTAGAAGCTT TAGAAAAGGA ACATCCTGAA GGTATTTATC TTTCTGGTGC ACTATTAATG 2640  
 5 TTTGCTGGCT CAAAATCATA TTAATTATAT GGTGCGTCTT CTAATGAATT TAGAGATTTT 2700  
 TTACCAAATC ATCATATGCA GTATACGATG ATGAAGTATG CACGTGAACA TGGTGCAACA 2760  
 ACTTACGATT TCGGTGGTAC AGATAATGAT CCAGATAAAG ACTCAGAACA TTATGGATTA 2820  
 10 TGGGCATTTA AAAAAGTGTG GGGAAACATAC TTAAGTGAAA AGATTGGTGA ATTTGATTAT 2880  
 GTATTGAATC ACCCATTGTA CCAATTAATT GAGCAAGTTA AACCGCGTTT AACAAAAGCT 2940  
 AAAATTAAAA TATCTCGTAA ATTAAAACGA AAATAGATTA ACGACTGAAA TCTGAACGCT 3000  
 15 CATAAGACTG TCATTTGCGT TCAGATTTTT TTACACAATA TAGAATGGTT GAGTAAAATA 3060  
 TTTTGAATA TAGTGAAAGA GGGGAAGTA CTGTGATAAA AAAGCTATTA CAATTTTCTT 3120  
 TAGGGAATAA GTTTGCTATC TTTTAAATGG TTGTTTTAGT TGTCTGGGC GGTGTATATG 3180  
 20 CGAGTGCTAA ATTGAAATTA GAATTACTAC CAAATGTACA AAATCCAGTT ATTTCACTTA 3240  
 CAACAACAAT GCCGGGTGCA ACGCCACAAA GTACCCAAGA TGAAATAAGT AGTAAAATTG 3300  
 25 ACAATCAAGT AAGATCATTG GCATATGTGA AAAATGTTAA AACGCAATCC ATACAAAATG 3360  
 CTTCAATTGT AACAGTTGAA TATGAAAATA ATACAGATAT GGATAAAGCA GAAGAACAGC 3420  
 TTAAAAAGA AATCGATAAA ATTAAATTTA AAGATGAAGT TGGTCAACCA GAATTAAGAC 3480  
 30 GTAATTCGAT GGATGCTTTT CCGGTTTTAG CATATTCATT TTCAAATAAA GAGAATGACT 3540  
 TGAAAAAAGT AACGAAAGTA CTGAATGAAC AATTAATACC AAAATTGCAA ACGGTAGATG 3600  
 GTGTGCAAAA TGCGCAATTA AATGGGCAGA CGAACCGTGA AATCACCCTT AAATTTAAGC 3660  
 35 AAAATGAACT TGAAAAATAT GGGTTGACTG CTGATGATGT AGAAAACTAT CTAAAAACGG 3720  
 CAACAGAAGC AACGCCACTT GGATTGTTCC AATTTGGTGA TAAAGATAAT CAATTGTTGT 3780  
 TGATGGTCAA TATCAATCTG TTGATGCTTT TAAAAACATA AATATTCAT TAACGTGGCA 3840  
 40 GGAGGACCAA GGGCATCTCA TCCCAAAGTG ACCATAAACC AAATTCAGCC ATGTCAGACG 3900  
 TTATCAGGCA TCACCACAGC AAATTCAAAG CGTCAGnCC AATATATAGT GGATGCCGCA 3960  
 nGAACTAGGG GTTTAGCGnT ATCAGTGGTG TGGCGACTCT ATTCTAAACG AT 4012  
 45

(2) INFORMATION FOR SEQ ID NO: 48:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7778 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear



(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

	CAATATAGGT CGCCGAGTTT CAACTaCATC AACTGGTTCA GTTACATTAG ATAATGCGCT	60
5	AGGTGTAGGT GGCTATCCTA AAGGACGAAT TATTGAAATT TATGGTCCTG AAAGTTCTGG	120
	TAAGACAACA GTAGCGCTTC ACGCTATTGC TGAAGTACAA AGTAATGGCG GGGTGGCAGC	180
	ATTTATCGAT GCTGAACATG CTTTAGATCC AGAATATGCT CAAGCATTAG GCGTAGATAT	240
10	CGATAATTTA TATTTATCGC AACCGGATCA TGGTGAACAA GGTCTTGAAA TCGCCGAAGC	300
	ATTTGTTAGA AGTGGTGCAG TTGATATTGT AGTTGTAGAC TCAGTTGCTG CTTTAAACACC	360
	TAAAGCTGAA ATTGAAGGAG AAATGGGAGA CACTCACGTT GGTTTACAAG CTCGTTTAAT	420
15	GTCACAAGCG TTACGTAAAC TTTCAGGTGC TATTCTAAA TCAAATACAA CTGCTATTTT	480
	CATCAACCAA ATTCGTGAAA AAGTTGGTGT TATGTTCCGT AATCCAGAGA CTACACCAGG	540
	TGGACGTGCA TTAATAATTCT ATAGTTCAGT AAGACTAGAA GTACGTCGTG CAGAACAGCT	600
20	TAAACAAGGA CAAGAAATTG TAGGTAATAG AACTAAATT AAAGTCGTTA AAAATAAAGT	660
	GGCACCACCA TTTAGAGTAG CTGAAGTTGA TATTATGTAT GGACAAGGTA TTTCTAAAGA	720
25	GGGTGAACCTT ATTGATTTAG GTGTTGAAAA CGACATCGTT GaTAAATCAG GAGCATGGTA	780
	TTCTTACAAT GGCGAACGAA TGGGTCAAGG TAAGGAAAAT GTTAAATGT ACTTGAAAGA	840
	AAATCCACAA ATTAAAGAAG AAATTGATCG TAAATTGAGA GAAAAATTAG GTATATCTGA	900
30	TGGTGATGTT GAAGAAACAG AAGATGCACC AAAGTCATTA TTTGACGAAG AATAGTACAC	960
	AAATTATAT CTATAGTTAA ACTTAGCAAA TATCCTTATA GGATTGATTG AAAGTGATAT	1020
	TCATCTCATA AAGCTAGAAT AATATCTAAC TTTATGGGAT ACACTACAAA TCGAGACTAT	1080
35	AAGGTTTTTT ATTTTATTTA TTATTACATT ATCAATAGTT TTATAATCGA GCTTCAAAAC	1140
	TTTAGAAAAT AGTAGAAATA GCATTCAATA TAGTGCAAAA GTGCAAATTG ATAACCTGAC	1200
	ACTTATCTCC TATAAACCGT ACAATTAATT TGTATGATTT ATATATAATT TCATAAAGTC	1260
40	ATATTGAATT TCATATAAAG AGCAAACCCCT AGAAAAGGAG GTGTTTGTGT GAATTTATTA	1320
	AGCCTCCTAC TCATTTTGCT GGGGATCATT CTAGGAGTTG TTGGAGGGTA TGTGTTGCC	1380
	CGAAATTTGT TGCTTCAAAA GCAATCACAA GCTAGACAAA CTGCCGAAGA TATTGTAAAT	1440
45	CAAGCACATA AAGAAGCTGA CAATATCAAA AAAGAGAAAT TACTTGAGGC AAAAGAAGAA	1500
	AACCAAATCC TAAGAGAACA AACTGAAGCA GAACTACGAG AAAGACGTAG CGAACTTCAA	1560
	AGACAAGAAA CCCGACTTCT TCAAAAAGAA GAAAACCTAG AGCGCAAATC TGATCTATTA	1620
50	GATAAAAAAG ATGAGATTTT AGAGCAAAAA GAATCAAAAA TTGAAGAAAA ACAACAACAA	1680

55

	CGCATCTCCG GTCTCACTCA AGAAGAAGCT ATTAATGAGC AACTTCAAAG AGTAGAGGAA	1820
	GAAGTGTAC AAGATATTGC AGTACTTGTT AAAGAAAAAG AAAAAGAAGC TAAAGAAAAA	1860
5	GTGATAAAA CAGCAAAAGA ATTATTAGCT ACAGCAGTAC AAAGATTAGC AGCAGATCAC	1920
	ACAAGTGAA CAACGGTATC AGTAGTTAAC TTACCTAATG ATGAGATGAA AGGTGGAATC	1980
	ATTGGACGAG AAGGACGAAA CATCCGCACA CTTGAACTT TAACTGGCAT TGATTTAATT	2040
10	ATTGATGACA CACCAGAAGC GGTTATATTA TCTGGTTTTG ATCCAATAAG AAGAGAAATT	2100
	GCTAGAACAG CACTTGTTAA CTTAGTATCT GATGGACGTA TTCATCCAGG TAGAATTGAA	2160
15	GATATGGTCG AAAAAGCTAG AAAAGAAGTA GACGATATTA TTAGAGAAGC AGGTGAACAA	2220
	GCTACATTTG AAGTGAACGC ACATAATATG CATCCTGACT TAGTAAAAAT TGTAGGGCGT	2280
	TTAACTATC GTACGAGTTA CGGTCAAAAT GTACTTAAAC ATTCAATTGA AGTTGCGCAT	2340
20	CTTGCTAGTA TGTTAGCTGC TGAGCTAGGC GAAGATGAGA CATTAGCGAA ACGAGCTGGA	2400
	CTTTTACATG ATGTTGGTAA AGCAATTGAT CATGAAGTAG AAGGTAGTCA TGTGAAATC	2460
	GGTGTAGAAT TAGCGAAAAA ATATGGTGAA AATGAAACAG TTATTAATGC AATCCATTCT	2520
25	CATCATGGTG ATGTTGAACC TACATCTATT ATATCTATCC TTGTTGCTGC TGCAGATGCA	2580
	TTGTCTGCGG CTCGTCCAGG TGCAAGAAAA GAAACATTAG AGAATTATAT TCGTCGATTA	2640
	GAACGTTAG AACCGTTATC AGAAAGTTAT GATGGTGTAG AAAAAGCATT TGCATTTCAG	2700
30	GCAGGTAGAG AAATCCGAGT GATTGTATCT CCTGAAGAAA TTGATGATTT AAAATCTTAT	2760
	CGATTGGCTA GAGATATTAA AAATCAGATT GAAGATGAAT TACAATATCC TGGTCATATC	2820
	AAGGTGACAG TTGTTGAGA GACTAGAGCA GTAGAATATG CGAAATAATT TTTGTCTCCC	2880
35	TCACAAATTA GTGAGGGAGC TTTTITAAGT TGTAGTCTTA AtCTAGTTAG ACAGCACTTT	2940
	ATCGTAATA ACTATATTAA ACAGTAGTTA TTTGAAAGTA AGACGGACCT TATATTAAAT	3000
	AAGAAGTTAT TGCTTTTAAT AAAATGTTT TAGGCTTCGT AATTACTATA TTTATATTAT	3060
40	GTAAACCTAT AAAGATGATT GGTTTTCTAT CCAATAAAAA AGAAGAGAAG ATGTAACACA	3120
	TCTTCTCTTC yGCAATATTA ATTAGGATTT ATTCTAAGT TGAGTTATTT TAATTGTAAA	3180
	TCTGTTTTCT TTAATTCITT TATAACTTCT GCAGTATCAT AACAATTTGT TGCAATTGTT	3240
45	GAATATCTCT CTGCTAAACG ATATGCATTA ATGTAAAGCT TTAAACTTTC TTTAGCTATA	3300
	TCCTCTGCAT CTTGGAATTT TGATGGGTTA GACATAACCA CTAATCTGC AAATTTTTCT	3360
50	GGATCAATAT TAATAGACAT GTATTTATTT ACAACTCCTA TTTATTTTGA TGTCTTAATA	3420
	CTAACATATT GAAGTTTTCA GACAAAGTAA TGTCTCTCTA TAATTGAAGA AAAATAATTC	3480

55

	GGATGAACAA AACATGAGAA TAATGTTTAT AGGGGATATC GTAGGTAAAA TTGGACGAGA	3600
	CGCAATTGAA ACGTACATAC CTCAACTGAA GCAAAAGTAT AAACCAACAG TTACAATTGT	3660
5	AAATGCTGAA AATGCAGCAC ATGGTAAAGG TTTGACTGAA AAAATATATA AACAATTACT	3720
	AAGAAATGGT GTAGATTTCA TGACTATGGG TAATCACACA TATGGTCAAC GTGAAATTTA	3780
	TGATTTTATA GATGAAGCAA AACGACTAGT AAGACCAGCG AATTTTCCGG ATGAAGCGCC	3840
10	GGGAATTGGT ATGAGATTTA TACAAATTAA TGATATTTAA CTTGCAGTTA TTAATCTGCA	3900
	AGGAAGAGCG TTTATGCCAG ATATTGATGA TCCTTTTAAA AAGGCAGATC AATTAGTCAA	3960
	GGAAGCACAA GAACAAACTC CGTTTATATT TGTTGATTTT CATGCAGAAA CAACTTCTGA	4020
15	AAAGTATGCA ATGGGATGGC ATTTAGATGG TAGASTAGCG CTGTTGTTGG AACGCATACA	4080
	CACATTCAA CAGCAGATGA ACGTATTTTA CCAAAGGGGA CAGGGTATAT AACGGATGTT	4140
20	GGTATGACAG GTTTTTATGA TGGCATTTTA GGAATAAATA AAACAGAGGT AATTGAGCGT	4200
	TTTATCACTA GTTTGCCACA AAGACATGTT GTTCCAAATG AAGGTAGAAG TGTATTATCT	4260
	GGTGTGTGTA TTGATTTAGA CAAAGAAGGT AAAACAAAGC ACATCGAACG TATATTGATA	4320
25	AATGATGACC ATCCATTTTC AACATTTTAA AATTACGTAA GTAAACATTC GAATTGGACC	4380
	CTATCGTCCA TTAGTATGAA TTTAATATAG TACCACTGTT TACATAGTAA ATCGGTGGTT	4440
	CTTTTTGTGA TCATTTAATA TGAAATATAT CCATAGGAGG CATATACTA TGAAACCACA	4500
30	ATTATCGTGG AAAGTTGGCG GTCAACAAGG CGAAGGTATT GAATCAACTG GGGAAATCTT	4560
	CGCTACGGCT ATGAATAGAA AAGGATATTA TTTATATGGA TATAGACATT TTTCAAGTCG	4620
	TATCAAAGGT GGACATACGA ATAATAAAAT TAGAGTTTCT ACGACGCCTG TTCATGCAAT	4680
35	TAGTGATGAT TTAGATATTT TGATTGCATT TGACCAAGAA ACAATTGATG TTAACCATCA	4740
	TGAAATGAGA GAAGACAGTA TTATTTTATC TGATGCCAAG GCTAAACCTG TGAAACCAGA	4800
	AGGATGTCAT GCACAGCTTA TTGAATTACC TTTTACAGCA ACCGCTAAAG AATTAGGTAC	4860
40	AGCATTAATG AAAACATGG TTGCAATAGG TGCTACTAGC GCATTGATGA ATTTGAATAC	4920
	AAATACATTT GAAGAACTTA TTAATAATAT GTTTTCTAAA AAAGGTGACA AGGTAGTTGA	4980
45	AGTCAATATC CAAGCATTA ACGAAGGTTA TCAATTAATG CAATCTCGCT TACCTGAAAT	5040
	CTACGGGGAC TTTGAATTAG AGTCAACAGA TGCACTACCA CATCTATATA TGATTGGTAA	5100
	CGATGCCATT GGATTAGGTG CAATTGCTGC AGGTTACAAA TTTATGGCGG CATATCCTAT	5160
50	TACACCTGCG TCTGAAGTTA TGAATATAT GATTGCCAAT ATATCTAAAG TAAACGGAGC	5220
	GGTTATTCAA ACAGAAGATG AAATTGCTGC TGTAACATATG GCTATTGGTG CAAATTATGG	5280

55

	TGGATTATCT	GGTATGACTG	AAACGCCATT	AGTCATTATT	AATACCCAAC	GAGGTGGACC	5400
5	TTCTACTGGA	TTACCTACGA	AACAAGAACA	GTCAGATTTA	ATGCAAATGA	TTTATGGTAC	5460
	ACATGGTGAT	ATTCCAAAAA	TTGTTGTAGC	ACCAACAGAT	GCAGAAGATG	CATTTTATTT	5520
	AACTATGGAA	GCATTTAATT	TAGCAGAACA	ATATCAATGC	CCTGTTATAG	TTCTAAGTGA	5580
10	TTTGCAATTA	TCTTTAGGTA	AACAACTGT	TGAAAAATTA	GATTATAATC	GTATTGAAAT	5640
	TAAACGTGGT	GAAATCATT	AATCTGATAT	TGAACGTGAA	GAAGATGATA	AAGGTTATTT	5700
	CAAGCGTTAT	GCGTtAACAT	CCGATGGTGT	TTCTCCTAGA	CCTATCCCCG	GTGTTAAAGG	5760
15	AGGTATTCAT	CATATAACTG	GTGTGGAaCa	CAATGAAGAA	GGTAAACCTA	GTGAATCTGC	5820
	GTCAAATAGA	CAACAACAAA	TGGAAAAACG	AATGCGTAAA	ATTGAGCAGT	TACTAATTGA	5880
20	ATCGCCAGTA	GAAGCTAACT	TACAACATGA	GGATGCAGAT	ATTCTTTATA	TCGGTTTTAT	5940
	TTCTACAAAA	GGTGCAATTC	AAGAAGGTAG	TAACCGTTTG	AATCAACAAG	GCATAAAAGT	6000
	TAACACTATA	CAAATTAGAC	AATTGCATCC	ATTCCCAACA	AGCGTTATTC	AAGATGCAGT	6060
25	TAATAAAGCG	AAGAAAGTCG	TTGTAGTGA	GCACAATTAT	CAAGGACAAT	TGGCTAGTAT	6120
	TATAAAAAATG	AATGTCAATA	TTCATGATAA	GATTGAAAAT	TATACAAAGT	ATGATGGGAC	6180
	ACCTTTCCTA	CCACATGAAA	TCGAAGAAAA	AGGCAAAATA	ATTGCTACTG	AAATAAAGGA	6240
30	GATGGTATAG	ATGGCGACAT	TTAAAGATTT	TAGAAATAAT	GTTAAGCCTA	ACTGGTGCCC	6300
	CGGATGTGGC	GATTTCTCAG	TACAAGCTGC	AATTCAAAAA	GCAGCCGCAA	ATATAGGGTT	6360
	AGAACCTGAA	GAAGTAGCTA	TCATCACCGG	TATAGGATGT	TCTGGCCGTC	TTTCAGGATA	6420
35	TATTAATTCT	TATGGCGTTC	ATTCTATTCA	CGGACGTGCA	TTACCTTTAG	CTCAAGGTGT	6480
	AAAAATGGCG	AATAAAGATT	TAAGTGTAT	TGCATCGGGA	GGAGATGGTG	ATGGTTATGC	6540
40	TATAGGTATG	GGGCATACAA	TCCATGCTTT	AAGAAGAAAT	ATGAACATGA	CGTATATAGT	6600
	CATGGATAAT	CAAATTTATG	GTTTGACAAA	GGGACAAACA	TCGCCGTCAT	CAGCAGTAGG	6660
	ATTTGTTACT	AAAACAACGC	CAAAAGGTAA	TATAGAAAAA	AATGTTGCGC	CTTTAGAATT	6720
45	AGTATTATCA	TCTGGTGCCA	CATTTGTAGC	CCAAGGTTTT	TCAAGCGATA	TTAAAGGATT	6780
	AACAAAACCTA	ATTGAAGATG	cAATTAATCA	TGATGGATTT	TCATTCTGTTA	ATGTCTTTTC	6840
50	ACCATGTGTG	ACTTATAATA	AAATTAACAC	ATACGATTGG	TTTaAAGAAC	ATTTAACAAG	6900
	TGTTGATGAc	ATTGAAAAAT	ATGATTCTAC	AGATAAACAA	TTAGCGACTA	AAACTGTTAT	6960
	TGAACATGAA	TCTTTAGTAA	CTGGTATTGT	TTATCaAGAT	AAAGAAACAC	CATCATATGA	7020
55	ATCtCAAATT	AAAGAGTTAG	ATGATmCACC	ACTTGCTAAA	AGAGATATCa	AAATTaCTGA	7080

TGTATTTATA ACAGATCCAT TTATGCTACT CAGTTTTTTA CTATTACAAA AAATAAAGGA 7200  
 GTTTTTAAAA ATGAAAGACA CATTAAATGAG TATACAAATA ATTCCTAAAA CACCAAACAA 7260  
 5 TGACAATGTT ATACCTTACG TAGACGAGGC GATTAAAATA ATTGACGAAT CTGGTTTGCA 7320  
 TTTTAGAGTA GGTCCGTTAG AAACGACAGT ACAAGGAAAT ATGAATGAAT GTTTAATTTT 7380  
 10 AATACAATCA TTAAATGAAC GAATGGTGGA ACTTGAATGT CCAAGTATTA TTAGCCAAGT 7440  
 TAAGTTTAT CATGTGCCAG ATGGCATCAC TATTGAACT TTAAGTAAA AATATGATGA 7500  
 ATAACATTAA AAGTGAAGTA AACTGGATT GAATTGGCTT GTTAGAGATG ACGTATAACT 7560  
 15 TTAAGTGT TTGCACTTA TAGTTAAAT TAATATAATT ATTAAATGAT ACGGGCAAAT 7620  
 AGAAAGGATT TTGTAAAGTG AACGAAGAAC AAAGAAAAGC AAGTTCTGTA GATGTTTTAG 7680  
 CTGAGAGAGA TAAGAAAGCA GAAAAAGATT ATAGTAAATA TTTTGAACAT GTTTATCAGC 7740  
 20 CGCCTAATTT AAAAGCAAGC GCAAAAAAG AGGTnAAA 7778

## (2) INFORMATION FOR SEQ ID NO: 49:

25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1128 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

AGATGAAGTT GTTACgAAAA TTGCGTACGC TGTTTCAGAA CATGTCAAAA TAGAAACAGG 60  
 35 TAATCCATT TTTCAAACAT CACATAGTGG TTGTGCGACG GGCGGATCCT GTAATTGTTC 120  
 ATTATAAAAA ACATCGAGTC AGAAAAAGGT GGTTATTGAA CCACTAACTA GCATCTGACT 180  
 CGATGTTTTT ATTTATTGGG GATTGTTTGT TTGAATTGTT GTGCTAAATC TGGTCGATCT 240  
 40 GTCACAATCG TGTGTGCACC TTTTGGTAT AAATCATTCA TCAGATTTAT ACTATTTACG 300  
 CCATAATAGC CTGGAATGAT ATTCATATCA TTAAACCAT TGATAAAACG AGATGAAGTC 360  
 45 AAATCAATGC CTTTAAAATG AGTAGGCATT TGGAACGTTT GTGCTAATGG TTGGTAGTAC 420  
 CTACCACCTA ATAAATGATA TTTTAAAAAT GCTTCTGTAA CTTCTGTG GCTAGCACCA 480  
 ATTGCGACGG ATCCTGTGC AATTTTATTA AAACGAACGA TTTGTTCTTT ATAAAACTT 540  
 50 GTCACAAGAA CGCGGTCAAA TGCTTGATTT TCTGCAATTG TATCAAACAT AATTTGTGGT 600  
 GCGATTGAGC CTTCATAGGA TTCAGGAGCA TCTTTTAAGT CTACGTTTAT ATACATATCA 660  
 55 GGATATTGCT TCAGCAACTC ATCGAAGGTT AGTATAGCTG TGTGTGCATG ACCACGATAT 720

AATGTATGGG CACTAACTTT TCCAGAGCCG TTCGTCGTTT TATCAACAGT TCGGTCATGA 840  
 AAAACGATAA GCTGTTGATC TTTTGTGAGT CTCACATCTG TTTCAAAGCC ATCAACGCCT 900  
 5 AATTGTTTAG CATAGTCAAA TGCAAGTTGC GTTTGCTCTG GTCTTAAAGC CATACCACCG 960  
 CGATGCGCAA ATATATATGG TGCATTGCCT TTGAAAAAAG CAGGGATGGT TTGCTTTTTTA 1020  
 10 GTAATCACTT TATTTTTTATT GATCATTAAT AGACTACTTA AAAATCCAGC ACCGACTAGT 1080  
 ACCGCATTTA AAATGTTTCT GTTTACHTTT TTCATAAAAA ATTCCTCC 1128

(2) INFORMATION FOR SEQ ID NO: 50:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6252 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

CAAGCAAACA ATCGTCGATA AAATTGCTAA AATAATAAAA GTAATTCGAA CTTTCATCAT 60  
 25 GATCATCCTT TGTATTATAGA GTCAATATAA GTATGGAATA TGTTAGGTAT ATAGTCAAAT 120  
 GCGTCAACTA ATGGGAATTT TGGCATAGAT AGAGAATTTA AGGCAATTAA AAAGGCATCA 180  
 30 AACAGTAATA TGCTGCTTGA TGCCCAAATG ATGACTTTAG CTAAATTGAT TAGTCACTTT 240  
 TAAAGATAAA GAATTGTCAT GAATTAAAAC TCATGTAATG ATGTGTTACA TTTCGCAATG 300  
 ATGGCTTTCA GTTATTTATC GATAACATCA CTCTTGATAC CTTTAGATTT TAAGAAATCT 360  
 35 TTAATTTTAT CTTGTTGCTT TTTATTAACA TCACCGGCAT ATTTTGTGTTG CACGTCGACA 420  
 ACATTGATTT TATTTTGCGG TTGATAGCTA AGCTTTTCAA TATCTTCATC AACATTGGCG 480  
 ATTGTACTAT TTAAAGCTTT GAAGTAATTC ATCATTAATT CAACGGGTTT CTTATATTCT 540  
 40 TTAGGAATAT TGTTTTCAGT GACAAATTTT TTGAAATGCA AATCGTTTTT AACAGCTAAG 600  
 TTAGATAAGT GGCTAAGTGT TTCTGCTTGT TTTTCAGTCA CTTTGTGTTG ACTGTCAATT 660  
 45 TGTTTATCTA GTTTATGTTG CATAATATAT TTGTTATCAA GTATATCGCT ATTTACAGAC 720  
 AAATACTTTT CTATAGCTTG CTTTCTCTCT GCATCACTAA TATCACTATT TTTCTTATCT 780  
 GAGTTAAAGA TATCTTTTGT TCTTAATTTT TTAGCGCTTT TAGGTGCATG GATGCCAGTA 840  
 50 CTTGTATGAT GATCTTCGTT ATCAGATTGA TCGGACGCGC AACCTGTAAG AATTAATGTC 900  
 GATGCTAAAA ATGTACTTAG TAGTAATCTC TTTTTCATAA TGTAATATAA CTCCTTAGTT 960  
 55 TATCTTTAAT TGAAAAAATA TGTATTCATG TTTAATAGAG TAACATTGAA TTAGTTTGGA 1020

	TCTATCAATA ATGCATCATT TTGGACGTTG TTAAGGATAG CTTTATCTAT AAATAACTGC	1140
	ATAATTGGTT GTACTAATTT AGACGTAGGT ATCGTACGTA AAAGCATAAT AATTTTCGTTT	1200
5	ACATACTTTT CTTTCTCAAT ATCATTCTTC ATATTGATTT GTTTGCGAGA GGTACATACT	1260
	TTAAGCATT A TCGCACATCT CGTTGTATAT ATTAAGTTTA TCATAACATG ATTTTATGTC	1320
10	GGGATAAAAA AATAACAGCA TCTTAACAAA TGTAAGATAC TGTCAGTGAA ATGAATGAAA	1380
	CTTTAGTTTC TGaTAATATA GTCAAAGGCA TTTAATGCTG CATTTGCACC AGCGCCCAT	1440
	GAAATGATAA TTTGTTTGTT CTTCTGATCT GTGACATCGC CAGCAGCAAA TATTCCAGGA	1500
15	ACATTCGTAT TATTGTTACG ATCAATCACA ATTTACCAC GTTCGTTTAA TTCAACAGCA	1560
	TCGTTTAACC ATGATGTGTT TGGAAGTAAA CCAATTGAA CAAAGATACC ATCTAAGTTA	1620
	AGTAGATGTT CTTCCGCGGT GTTCATGTCT TCGTAACGTA TACCTGTAAC ATGGTCTTCT	1680
20	CCGACAACTT CAGTAGTTTT GGCATTGTGTT TTGATATCAA CATTTGATAA AGAACGTAAA	1740
	CGATCTTGTA ACACGTTGTC TGCTTTTAAT TCGCTAGCGA ATTCGAATAA TGTAACATGA	1800
25	TTAACGATAC CAGCAAGGTC AATTGCTGCT TCAACCCAG AGTTACCGCC ACCGATAACT	1860
	GCTACGTCTT TATTTTCAAA TAGAGGTCCG TCACAGTGAG GGCAGAATGC AACACCTTTA	1920
	TTAATCAATT GCTCTTCACC TGGAATGTTT AGCTTACGCC AACCTGCACC AGTAGCAATA	1980
30	ATGACTGTTT TACTTTCTAA GACAGCACCG TTTTCTAACG TAACTTTAAT TGCTTCGTCA	2040
	GTCTTTTCGA TATCTGTAGC ACGTATACCT GTCAATGCAT CAATGTCATA TTGATCAATG	2100
	TGCGCTGCTA AGTTAGAAGA AAATTCAGAA CCAGTTGTTT CTTTAACAGT AATGAAGTTC	2160
35	TCAATACCAG CAGTATCATT AACTTGGCCA CCGATACGAT CAGCAACTAT ACCAGTACGT	2220
	AAACCTTTAC GTGCTGTGTA AATCGTGCA CTACCACTAG CAGGACCACC ACCAACGATT	2280
40	AAGACATCAT AAGGTTCTTT ATTTTCAAAC TCAGATGCAT CTGCCGTACT GCCTAGTTTC	2340
	GAAAGAATAT CTTGGATTGT CATACGACCA TTGCCAAATT CTTGCCCAT TAAAAAGACA	2400
	GCAGGGACTG CCATGATGTT TTCAGATTCT TCACGGAACA CTGCACCATC AATCATAGAA	2460
45	TGCGTGATGT TAGGGTTGAT CACACTCATT AAGTTAAGTG CTTGAACGAC ATCAGGACAT	2520
	TTTTGACACG TTAAACTAAT GAATGTTTCA AAATGGAATG AACCTTCTAA TTTTTTAATT	2580
	TGGTCAATGA TTGACTGTTT TTCTTTAGGT GCACGACCAC TAACCTGTAA AATTGCTAAA	2640
50	ACAAGTGAGT TAAACTCGTG ACCTAATGGA ATACCTGCAA ATGTTACACC TGTTTCTTCG	2700
	CCAGGACGAT TGA CTGAGAA ACTTGGTGTA CGTTTTAAAG ATTTTTCAGA AAGAGATAGT	2760
55	CTAGGTGACA TATCAGTAAT TTCTGTCAAC AAATCTTTAA GTTCTTTGGA TTTATCATCT	2820

	TGTTGTTTTA AATCAGCAIT AAGCATGGTT GTAATGCCTC CTTAGATTTT ACCTACTAAA	2940
5	TCTAAACCAG GTTGCAATGT TTTAGCGCCT TCTTCCCATT TAGCTGGGCA TACTTCGCCA	3000
	GGGTTTTTAC GAACATATTG AGCTGCTTTG ATTTTGTGAG CTAATGTACT AGCGTCACGG	3060
	CCAATTCCGT CAGCGTTAAT TTCAGATGCT TGTACAACAC CGTCTGGGTC GATAATGAAT	3120
10	GTACCACGTT GAGCTAAACC AGTAGCTTCA TCTAATACAT CAAAATTACG AGTGATTGTT	3180
	TGTGATGGGT CACCAATCAT AGTGTAAGTG ATTTTGCTAA TTGCATCTGA ATGGTCATGC	3240
	CATGCTTTGT GTACGAAGTG AGTATCAGTT GATACTGAGA ATACATTTAC GCCTAATTTT	3300
15	TGTAATCTT CATATTGGTT TTGTAAGTCT TCTAATTCAG TTGGACAAAC GAATGAGAAG	3360
	TCAGCAGGAT AGAAGCATAC TACGCTCCAA GAACCTTTTA AATCTTCTTG TGTAACTTCT	3420
	TTAAATTGAT CTTTTTTTGG ATCGAAATCT TGCGCTGTAA ATGGTAAGAT TTCTTTGTTA	3480
20	ATTAATGACA TAAATATCTT CCTCCTAAGA ATTTAAGTAT GAATTAGAAC TATCAATTGA	3540
	TTGCGCTTAA TTATAATAAT TCTAATCTCT TAGTTAGCAT TATTACATTT TGATCCAGAA	3600
25	TAGTCAACTG GATAACTTTG TAAAGTGAAT GATTACTTTT AAAATAAAGA AAGATAATAT	3660
	AAAGTGCTTT GATAATGGAT TTTGTAGTTG ATGATTTAAA AGGTTGTGTC TATATTTAAT	3720
	ATCTTGATTT TAATGTAAAA AATGTAAAAA AAGAAGATTT GTATTCTCAA CTAAGTCAAC	3780
30	CTTATTGATA ATGGTATGAG AATATTTGTT CGAGATGGAT GAAGGTAATG AGTGAGAAAC	3840
	TGGATTTTTA AAGTATGAGA CAATATTTTA AAAAGTTCAA TTATTAACCT ATAAGCAAAT	3900
	AATTGCTATA AAAAAGTTTG GACGTGTACA ATTGCAATAT GAAGATTTTA AATTAATTGT	3960
35	AAAGTATCGA GGAGTGGGTA ACGTGTGAGA ACATGTATAT AATCTTGTGA AAAAGCATCA	4020
	TTCTGTTAGA AAATTTAAGA ATAAACCTTT AAGTGAAGAC GTTGTTAAGA AATTGGTAGA	4080
40	AGCTGGACAA AGCGCTTCGA CGTCAAGTTT CCTGCAAGCA TACTCAATTA TTGGTATCGA	4140
	CGATGAGAAG ATTAAAGAAA ATTTACGAGA AGTTTCTGGA CAACCTTATG TTGTAGAAAA	4200
	TGGCTATTTA TCGTCTTTG TTATTGATTA TTATCGTCAT CATTAGTTG ATCAACATGC	4260
45	TGAAACTGAT ATGGAAAATG CATATGGTTC AACGGAAGGT TTGCTAGTAG GTGCAATCGA	4320
	TGCAGCATTG GTTGCCGAAA ATATTGCGGT AACTGCTGAA GATATGGGGT ATGGCATTGT	4380
	CTTTTTAGGA TCATTAAGAA ATGATGTTGA ACGCGTTCGA GAAATTTTAG ACTTACCTGA	4440
50	CTATGTCTTC CCGGTATTTG GTATGGCAGT AGGGGAACCC GCAGATGACG AAAATGGTGC	4500
	AGCCAAGCCA CGCTTACCAT TTGACCATGT CTTCCATCAT AATAAGTATC ATGCTGATAA	4560
55	GGAAACACAG TATGCACAAA TGGCAGATTA CGACCAGACA ATCAGCGAGT ACTATGATCA	4620



CAAAGCAAGA TTAGATATGT TAGAACAATT GCAAAAATCA GGCTTAATAC AGCGATAgCA 4740  
 AGATACCAAA ATAACCCGCC CCCCTCTAGC TTAATATGAT AAGTATAGCT AGAGGGGGCG 4800  
 5 GGTATTTCTT GCAATGAATT AGTGTGAAGT TAATGCAGCA TTATCATTTG AATCGAAAGT 4860  
 ATCTTTATCC CAATGTTTAG TTAACCTGGC GGTACCTGTA CCAGCTAGCA TTGAATCGTT 4920  
 CACGTTTAAT GCTGTTCTAC CCATGTCAAT CAATGGTTCA ACGGAGATGA GCACGCCGGC 4980  
 10 TAAAGCGACT GGCAAGTTTA ACGTTGACAA CACCAATATG GATGCAAATG TAGCCCCGCC 5040  
 ACCGACGCCA GCAACGCCGA ATGAACTAAT AATCAGGACA GCGATTAACG TTACAATAAA 5100  
 15 TTGTAAATCA ATTTCTACAT TAGCGACGGG TCGGACCATA ATTGCAAGCA TGGCAGGGTA 5160  
 AATGCCTGCA CAACCATTTT GTCCAATCGA CAATCCAAAT GTCGCAGCGA AATTGGCAAT 5220  
 ACCTTCTGGC ACGCCTAGAC GTCTTGTGTTG TGTGTTGTACA TTCAATGGTA AGGCACCCGC 5280  
 20 GCTTGAGCGT GATGTGAATG CAAAGATTAA TACTTCCAAA GTCTTTTAA CATAGCGAAT 5340  
 TGGGCTAATA CCTAACAGGC TTAATAAAT TAAGTGAATG ATATACATCG TAATTAATGC 5400  
 AGCGTACGAT GCGATTAAGA ATTTTCCTAA AGTCCAAATG GCGCCAAAGT CACTTGTCGA 5460  
 25 TAATGTGTTG GCCATAATTG CTAATACACC GTATGGCGTT AAACGTAAGA CGAACGTCAC 5520  
 AATCGCCATT ACTAGTGAAT AGATAGCGTC AATCGCACGC TTAAGCAATT CACCATGATC 5580  
 30 AGGTGTTTG CGTGTACGCG TAAATAAGCA AATCCTATAA ACGAAGCAAA TATCAGACA 5640  
 GCAATCGTGG aAGTTGCACG TTGTCCaGTG AAATCTAAGA ATGGATTTT AGGCAATAAT 5700  
 TCCAAAATTT GTTGTGGTAA CGTATGTGCT GTTAAATCTT TCGCTTGTT AGCAATTTG 5760  
 35 CTTCCACGTG CTGTTCAGC GTTACCAAGG TTAATGTTG ATGCATCTAA ACCAAACACC 5820  
 AAGGCATACA CAACACCAAC AATCGCAGCA ATGGTGACAG TGCCAATTAA AAAGATAAAA 5880  
 ATGAGACTAC CAATTTTAGC AAACTTTCT CCGATTGAA TTTTAGTGAA TGCAGCTACA 5940  
 40 ATAGAAATGA AAATTAAAGG CATAACAATC ATTTGCAACA ATGCAACGTA ACCTTGTCG 6000  
 ACAATGTTGA ACCAGTCACT TGTGATGTA ATAACATTCG AATGTGTGCC ATAAATAAGA 6060  
 TGCAATAACA CACCGAATAC TATACCAATC CCTAAAGCTG TAAACACACG TTTCGCAAAA 6120  
 45 GATATATGTT TCGGAGCCAT CATGTGCAAT ATTACGATGA AAATCACCAA TACAATAATA 6180  
 TTAATCAGTG TAAGAAAAGC ATTCATGAAC GTCACCTCTT AAATTTTGA ATATAATTCC 6240  
 50 GACTAGTATG CT 6252

(2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6730 base pairs

55

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

	ATCAAATcnc AAAATATTTA TTAATnAnAA GGGGATTATC CaTGtGAGAA ACAAAGTAAT	60
10	GCTCTTTTTT TACCTCTTGT GGGTTGAAAA aTGGATCATC AGAGATAGAC TTCTTCTTTT	120
	TCGAAGATGA CATTGATAC TTTAATCTTC TAAAACCATA ACTTGTCGCA TCAAAAATGC	180
	CTTCTGTAC AAGTAAAATC AAAAATATGC TAATAAAAAT AATTAATGAA ACATAAAACA	240
15	ATATATTTAA ATATGTAATG ATAGTATGGC TATTAAAAAG CCATATAATA AACGTTAATA	300
	TTGGCGTTAT TAGTGCCATT CCAAGCCATT TTTTCAACAT TTGATCACTC CCACTTATAG	360
	AAACTCTTA CGCATAGTTT ACATTAAAT CAGACATTGA GGAATGATTT TTTAATTCT	420
20	TCAGCTTTAT TGAAATTCTA AAATCAATCA TTCTTCATTA GTTTAAAGCA AAAAAATATT	480
	GATATATAGT AAATATTGTA TATATAATAT TAGTTAAGAT TTCaGAAAAAT TTTGAAGGGA	540
25	ATGGAATTT AGAAATCGGA ATTTGTTAGA GGAGGGGATT AGATGGGGAA ATATATTTTC	600
	AAACGATTTA TTTATATGCT TATTTCTTTA TTTATTATTA TTACAATTAC ATTTTCTTA	660
	ATGAAATTAA TGCCAGGTTT GCCATTTAAC GATGCTAAAT TAAATGCTGA ACAAAAAGAA	720
30	ATTTTAAATG AAAAAATATGG ATTAAATGAT CCTGtAGCTA CGCagTATTT ACATTATTTA	780
	AAAAATGTTG TTACAGGCGA TTTTGGTAAT TCATTCCAGT ATCATAATCA ACCTGTGTGG	840
	GATTTGATTA AACCGAGACT ACTACCTTCT TTTGAAATGG GTCTTACAGC AATGTTCaTC	900
35	GGTGTGATAC TGGGACTTAT TTTAGGTGTT GCAGCAGCTA CTAAACAAAA TTCTTGGGTT	960
	GACTATACAA CTACAGTTAT TTCAGTTATT GCAGTATCTG TACCATCTTT TGTACTTGCT	1020
40	GTACTTTTAC AATATGTATT TGCAGTTAAA TTAAGATGGT TCCCAGTAGC TGGATGGGAA	1080
	GGTTTTTCGA CCGCGGTATT ACCGTCACCT GCATTATCTG CAGCTGTTTT AGCAACTGTC	1140
	GCCAGATACA TAAGAGCAGA GATGATAGAG GTATTAAAGTT CAGACTATAT TTTATTAGCG	1200
45	AGAGCTAAAG GTAATTCGAC AATGCGTGTA CTTTTGGAC ATGCACTTAG AAATGCTTTA	1260
	ATTCCAATTA TTACAATTAT CGTTCCCATG TTAGCAAGTA TTTTAACAGG CACTTTAACA	1320
	ATTGAAAATA TTTTGGAGT TCCTGGATTA GGGGATCAAT TCGTACGTTT AATTACAACA	1380
50	AATGATTTCT CAGTAATCAT GGCAATCACA CTATTATTTA GCACACTGTT TATCGTTTCT	1440
	ATTTTTATTG TAGATATTTT GTACGGTGTG ATAGATCCAC GAATTCGTGT TCCaAGgAGG	1500
55	TAAAAAATAA TGGCTGAAAA TAAAAACAAT TTGTCGATTA ACGACGATCA TTCTAATGCA	1560

	TGAATCAGGA ACCTGAAATG CAACGAGAAA GCAAAAACCTT TTGGCAAGAT GCTTGGGCTC	1530
	AGTTAAAACG AAATAAGTTA GCTGTTGTCG GTATGATAGG TTAAATTATC ATTGTAATAT	1740
5	TTGCTTTTAT CGGTCCAGTT ATAAATAAAC ATGATTATGC TGAACAAAAT GTAGAACATA	1300
	GAAATCTTCC GGCAAAAATA CCTGTATTAG ACAAAGTTCC ATTTTTACCT TTTGATGGTA	1850
10	AAGATGCAGA TGGCAAGGAT GCTTATAAAG CAGCAAATGC TAAAGAAAAT TATTGGTTTG	1920
	GTAATGATCA GTTGGGTCGA GATTTATGGA CAAGAACATG GAAAGGTGCT CAAATTTTCAT	1980
	TGTTTATCGG TGTGTTGCA GCGATGTTAG ATATTTTTAT TGGTGTGTA TATGGTGCGA	2040
15	TTTCTGGATT CTTGGGTGGA CGTGTGATA CGATTATGCA ACGTATACTT GAAGTCATAG	2100
	CATCTATTCC GAATTTAATT GTCGTAATTT TATTTGTATT AATTTTTGAA CCATCCATTT	2160
	GGACAATTAT ATTGGCTATG TCTATCACAG GCTGGTTAGG CATGAGCAGA GTTGTACGTG	2220
20	GAGAATTTTT AAAATTAAAA AATCAAGAGT TTGTCATGGC TTCGAAAACA TTGGGGGCTT	2280
	CAAAATTCAA ATTGATATTT AAGCATATTT TACCTAATAC ATTAGGTGCT ATCGTGGTTA	2340
25	CATCAATGTT TACAGTACCT AGTGCTATTT TCTTCGAAGC ATTTTAAAGT TTCATTGGTA	2400
	TAGGTGTACC CGCACCTCAA ACATCGTTAG GGTCAATAGT AAATGATGGG CGCGCAATGT	2460
	TATTAATTTA TCCACATGAA TTATTTATAC CAGCAATGAT TTAAAGTTTA TTAATTCTAT	2520
30	TCTTTTACTT ATTTAGTGAT GGATTACGTG ATGCATTTGA TCCGAAAATG CGTAAATAAA	2580
	AAGGGGGCAT AGCATATGAC TGAAAGAATA TTAGAAGTAA ATGATTTGCA TGTTTCCTTT	2640
	GATATTACAG CAGGGGAAGT GCAGGCAGTG AGAGGCGTAG ATTTTTATTT GAACAAAGGG	2700
35	GAAACATTGG CAATTGTTGG TGAATCAGGT TCAGGTAAAT CTGTAACAAC AAAAGCAATT	2760
	ACAAAATTAT TCCAAGGGGA CACAGGAAGA ATTAATAAAG GAGAAATTTT ATTTTTAGGG	2820
40	GAAGATTTAG CAAAAAACC TGAAAATGAG TTGATTAAAT TACGTGGCAA AGATATTTCA	2880
	ATGATCTTTC AAGATCCAAT GACATCTTTA AACCCAACGA TGCAAATTGG TAAACAAGTC	2940
	ATGGAACCAT TAATTAAGCA CAAAATTAT AGTAAAGCAC AAGCTAAAAA GCGCGCATTG	3000
45	GAAATACTAA ATCTTGTAGG TTTACCAAAT GCAGAAAAAA GATTTAAAGC ATATCCTCAT	3060
	CAATTTTCAG GTGGACAAAG GCAAAGAATT GTTATTGCAA CCGCATTAGC TTGTGAACCT	3120
	AAAGTGCTCA TTGCTGATGA ACCAACGACT GCATTAGACG TAACGATGCA GGCACAAATT	3180
50	TTAGATTTAA TGAAAGAACT ACAACAAAAA ATCGATACAG CAATTATTTT TATAACGCAT	3240
	GATTTAGGGG TTGTTGCGAA TATTGCTGAT AGAGTGGCAG TTATGTATGG TGGTCAAATG	3300
55	GTTGAAACAG GAGATGTTAA CGAAATATTT TATGATCCAA AGCATCCATA TACATGGGGA	3360

	GGAGCGCCAC CTGATTTATT ACACCCACCT AAAGGTGATG CATTTGCGAG ACGTAGcAAT	3480
5	ATGCATTAGA TATTGATTTT AAAGTAGAAC CACCGTGGTT TAAAGTTTCA CCGACACATT	3540
	TTGTGAAATC TTGGTTATTA GACGCACGTG CACCAAAAGT TGAAC TACC GAGCTGGTAA	3600
	AACAACGTAT GAAACCGATG CCTAATAATT ATGAAAAACC ACTCAAGGTA GAAAGGGTGT	3660
10	CGTTCAATGA AAAATGATGA AGTGCTATTA TCTATTAAAA ATTTAAAGCA ATATTTTAAC	3720
	GCAGGAAAGA AAAACGAAGT GgaGCGATTG AAAATATTTT GTTTGATATA TACAAAGGGG	3780
	AAACATTAGG TTTAGTAGGA GAATCGGGGT GTGGTAAATC TACAAC TGGT AAATCAATTA	3840
15	TTAAACTTAA TGATATTACA AGTGGAGAAA TTTTGTATGA GGGTATTGAT ATACAAAAGA	3900
	TTGTAACG TAAAGATTTG CTTAAATTTA ATAAAAAGAT ACAGATGATT TTTCAAGACC	3960
	CATATGCGTC TTTAAATCCT AGGT TAAAAG TAATGGATAT AGTAGCTGAA GGTATTGATA	4020
20	TCCATCATTT AGCAACTGaT AAGCGTGACC GAAAAAACG TGTCTATGaT TTA CTTGaAA	4080
	CTGTTGGATT AAGTAAAGAA CATGCCAATC GCTATCCTCA TGAATTTTCA GGTGGaCAAC	4140
25	GCCAACGTAT TGGaATTGCC CGTGcATTAG CCGTTGaACC AGAATTCATT ATCGCGGACG	4200
	AACCAATATC GGCATTGGAT GTTTCAATCC AAGCTCAAGT AGTTAATTTA TTATTAAAAT	4260
	TACAACGTGA AAGAGGGATT ACGTTCCTAT TTATAGCTCA TGATCTATCA ATGGTGAAGT	4320
30	ATATTT CAGA TCGTATTGCA GTCATGCATT TTGGGAAAAT AGTTGAAATT GGACCGGCAG	4380
	AAGAAATTTA TCAAAATCCA TTACACGATT ATACTAAGTC TTTATTATCA GCCATTCCAC	4440
	AACCTGATCC TGAATCAGAA CGCAGTCGCA AACGATTTAG TTATATTGAT GATGAAGCAA	4500
35	ATAATCATTT AAGACAATTA CATGAAATTA GACCGAATCA CTTTGTCTTT AGTACTGAAG	4560
	AAGAAGCGGC ACAACTACGA GAAAATAAAT TGGTGACACA AAATTAAGGG GAAGGGGGAA	4620
	ATGcAATGAC GAGAAAATTT AGAACACTTA TTTTAATTTT GATTGCTACA ATTGCATTAA	4680
40	GTGGTTGTGC TAATGACGAT GGTATTTATT CAGATAAAGG TCAAGTATTC AGAAAAATTT	4740
	TGTCATCAGA CTTAACATCC CTTGATACAT CATTAAATAC GGATGAAATA TCTTCTGAAG	4800
45	TGAcTGCGCA AACATTCGAA GGT TTTATACA CATTAGGAAA AGGTGACAAA CCGGTGTTAG	4860
	GTGTTGCGAA AGCTTTTCCT GAAAAGAGTA AAGATGGTAA AACTTTAAAG GTTAAATTAA	4920
	GAAGCGATGC TAAATGGAGC AATGGTGACA AAGTGACTGC ACAAGACTTT GTTTATGCTT	4980
50	GGAGAAAAAC AGTTGACCCT AAAACAGGTT CTGAATTTGC ATACATTATG GGGGACATTA	5040
	AAAATGCGAG TGATATTAGT ACTGGTAAGA AACCTGTAGA GCAATTAGGT ATCAAAGCAT	5100
55	TAAATGATGA AACATTACAA ATTGAATTAG AAAAGCCGGT TCCATATATT AATCAATTAT	5160

ACGGTACGGC AGCTGATAGA GCGGTATACA ATGGTCCaTT TAAAGTTGAT GATTGGAAAC 5280  
 AAGAAGATAA AACCTTACTA TCTAAAAATC AGTATTATTG GGATAAAAAG AATGTAAAAT 5340  
 5 TAGATAAAGT GAATTATAAA GTTAATTAAAG ACTTACAAGC CGGTGCATCA TTGTATGATA 5400  
 CTGAATCAGT AGATGACGCA TTTATTACTG CAGATCAAGT AAATAAATAT AAAGACAACA 5460  
 10 AAGGATTAAA CTTTGTGTTA ACGACTGGGA CATTTTGTGT AAAAATGAAT GAAAAACAAT 5520  
 ATCCTGATTT TAAAAACAAA AATTTAAGAT TGsTATCGCA CAAGCAATAG ATAAAAAAGG 5580  
 ATACGTTGAT TCAGTGAAAA ACAATGGCTC AATTCCTTCC GATACACTAA CAGCCAAAGG 5640  
 15 AATTGCGAAA GCGCCTAATG GCAAAGATTA TGCGAGTACC ATGAATTCGC CTTTAAAATA 5700  
 TAATCCTAAA GAAGCAAGAG CACACTGGGA CAAAGCTAAA AAAGAGTTAG GTAAAAATGA 5760  
 AGTGACATTT TCAATGAACA CAGAAGATAC ACCAGATGCA AAAATATCTG CTGAATATAT 5820  
 20 CAAATCGCAA GTTGAGAAAA ATTTACCAGG AGTTACTTTG AAAATTAAGC AATTACCGTT 5880  
 TAAACAAAGA GTATCACTAG AACTGAGTAA CAATTTTGAA GCATCACTTA GTGGTTGGTC 5940  
 TGCAGATTAC CCTGATCCTA TGGCTTATTT AGAAACAATG ACCACAGGTA GCGCACAAAA 6000  
 25 TAATACAGAC TGGGGTAATA AAGAATATGA TCAATTACTT AAAGTAGCAA GAACCAAATT 6060  
 GGCACCTCAA CCGAACGAAC GATATGAAAA CTGAAAAAAA GCAGAAGAAA TGTTCTTAGG 6120  
 30 AGATGCACCG GTAGCACCAA TTTATCAAAA AGGTGTtGCA CATTTaACAA aTCCTCAAGT 6180  
 AAAAGGATTA ATTtACCATA AATTTGGTCC AAATAACTCA CTTAAACATG TATATATTGA 6240  
 TAAATCGATA GATAAAGAAA CAGGTAAGAA GAAAAAATAA TATGCTTTGT AAATTAGGCT 6300  
 35 GGAGACATAT CTCCAGTCTT TTTGTGTTGG ATAAAAaCTT TGGGAATAAA AATTTAAAAT 6360  
 AAGTCGTTTT TTAAATTACT GAAATTGATT AAATGCATAA ATAACTGAAT ATTCTAAAAA 6420  
 TAAAcTTGTA ATAATTTTTT CTATGAGTAA ACTAAAAAGA AAAAATTAGA TTGAAAGTAG 6480  
 40 GAGGCATATG TATGGGGAAG CTAATTAAAT ATATTTCAAT ACTTCTTATT GTCGTTTTAG 6540  
 TGTGAGTGC TTGCGGAAAA AGCAGTAATA AAGATGAAGG AGTAAAAGAT GCTACTAAAA 6600  
 45 CGGAAACCTC AAAACATAAA GGTGGTACCT TAAATGTAGC ATTAACAGCA CCGCCAAGTG 6660  
 GTGTTTATTC TTCGTTATTA AATAGTACAC ATGCAGATTC TGTAGTTGAG GGATATTTTA 6720  
 ACGAAAGCTT 6730

50 (2) INFORMATION FOR SEQ ID NO: 52:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 6482 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double

55

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

5	AATTTTGTGC ATTATTAAAA ACCTCGCTTT TAAAAGATTG AAAAGTAAAT GAGTGAAATT	60
	AAAGATTATG CACATTAAAA TCACGCCACA ATTTAATTGT GAAAAATATC ACAAATATAT	120
	TATAACACTA AATTTCCTAA AATTCAAAAG TGTGTTTTAT TGCAGAAAAC TTATAACATG	180
10	TGCACAAGTT ATAGTGAATT GCAAACGGAT TACTTTAGTC TTTTAAAAAC ATGAAGTATA	240
	ATTTGTATAG CAATAAATAT AAAAATGGGA GGCTATGTTT AATGAGCAAT ATGAATCAAA	300
	CAATTATGGA TGCATTTTAT TTCAGACATG CGACTAAGCA ATTCGATCCA CAAAAGAAAG	360
15	TTTCGAAAGA AGATTTTGAA ACAATATTAG AGTCAGGTAG ATTGTCTCCA AGTTCTCTTG	420
	GGTTAGAACC TTGGAAGTTT GTCGTGATTC AAGATCAAGC GTTACGTGAT GAATTAAGAG	480
	CGCACAGTTG GGGCGCAGCA AAACAATTAG ATACAGCGAG CCATTTTGTG CTAATTTTGTG	540
20	CGCGTAAAAA TGTAACGTCA AGATCACCGT ATGTACAACA TATGTTAAGA GATATTAAAA	600
	AATATGAGGC ACAAACGATT CCAGCTGTTG AACAAAAAAT CGATGCATTG CAAGCAGATT	660
25	TCCATATTTT TGATAATGAT CAAGCCTTGT ATGACTGGTC AAGTAAACAA ACGTATATTG	720
	CATTAGGCAA TATGATGACG ACAGCCGCAT TGTTAGGTAT TGATTCATGT CCGATGGAAG	780
	GTTTTAGTCT GGATACAGTG ACAGACATTT TAGCAAATAA AGGGATCTTA GATACTGAGC	840
30	AATTTGGTTT ATCAGTGATG GTCGCATTTG GCTACAGACA ACAAGAGCCA CCGAAAAATA	900
	AAACACGCCA AGCTTATGAA GATGTTATTG AATGGGTTGG ACCAAAAGAA TAAATAGAAT	960
	ACCGTATGTC TAAATATATA AAATTAAGAA GTTAGCAATA AAAAAGCCTG CGATTACATA	1020
35	AATGAATCGC AGGCTTTTGC GTGAAAAAAT TGTATTAATA AAGTATGGAT GATTATTTTT	1080
	CTGGTACAAG GTCAGTATTT GAATGAACTG TGATGTCAAA CCCTTCTGGT GCCGTAAATG	1140
	TATGTGTTGA GGCGTCGGGT TGATAAATAT CAACATGTGT TAATCCATAA CTTGTGGAAT	1200
40	TGTTTTGTCT TGCTTGATTG GATTGCCAAG TATTAGCAGC AATATGATGG TGATAATGAT	1260
	TCGTTGACAT AAATAGCGCA CGTGGAAAT CAGACACATG TTGGAATCCT AATTGTTCAA	1320
45	TGTAACATTG ATATGCTGCG TCTAAATCAT GTGTTTTTAA ATGTAAGTGT CCAATCATGC	1380
	CTTTTGCTGG CATTCCTTGC CAACCTTCAT CAGTACGATG TGTTAATAAG GTTTGGCTAT	1440
	CAACTTCTAA AGTATCCATT TTAACCTTGC CATTTTGCCA TTCCCATGAA GATGAAGGTC	1500
50	TATCGCGATA GACTTCAATA CCATTACCTT CGGGGTCGTT GAAATATAAA GCTTCACTTA	1560
	CTAAATGATC ACCAGCGCCG ATGCCCATAT TTTTTTGTGC CACGAAATAT AAGAAGTTAG	1620

55

	aAGTCTGACG GcCGTCTTCT AATAAATGTA ACGTTAGAGT ATGGcCACCA GTCCCAACAG	1740
	ATAATACGGT TGTATTATCG TCAGAACTTT TAACGGATAG TCCTAAAATG TTTTGTAAA	1800
5	ATGTTGTCAT TAAGTCTAAG TCTCTTACGT TCAGTACAAT GTTTGTCACT TGTGTTGCTG	1860
	TTTTATCGTG AAATGCCATT ATGCATCGCC TCTTTTTCTA TTTTCTATA AGTTAGTATA	1920
10	AAAAGTATAC CAGAAAAGAA AATGAATTGA TAGCATAAAG TTTGAAATGC AAAATAACTA	1980
	GTCGTTTTGC AATTTTAtAT TGATGCGAAC AAAAAAGCGA TGGTACAGTT GCACCATCGC	2040
	AAAATTTATT TAACCAAGAT ATACATCTTG ATATGAATCT TCTTTTTCTA ACATATGTTT	2100
15	GGCAATGAA CATGAGGCAA TAATTTTCAA ATTATTTTCT CGAGCGTGTT CAACAAC TGc	2160
	TTTAAGTAGT TTTTGGCAA CACCTTGACC ACCAAGTTCA TCAGATACGC CTGTATGATC	2220
	AATGTTAATT TCATTATTAT CCACAAAACG GTATGTGATT TCAGCTAAAG CATTATTTTC	2280
20	ATCATCACCA ATATAGAATT TGTCTCGCC TTGTTTGATT TCAAGGTTAC TCATACATAT	2340
	CAACTCCTAT CATGATTGAT TATAGTATTT CCTATTCTA TTTTAACTTA AACGAAGTCA	2400
	AAGGTGCATG ACAGTCATGT GACGACATTG CCACATCTAT GTAGTCGTTT TTATTAAGCA	2460
25	CAGTTTGAAA TGAAGATGAA AACACGTATC TTGACATTAA ATCTATTCAG CTATATAATT	2520
	TATCTCGAAA TCGAAATAAA ATAAAAAAGT TGGTGATCAT ATGGATCGAA CGAAACAATC	2580
30	TCTCAATGTT TTTGTGCGAA TGAATAGGGC GTTAGACACA TTAGAGCAAA TTACAAAAGA	2640
	AGACGTAAAG CGATATGGCT TAAATATTAC TGAATTTGCA GTGCTCGAGT TGCTTTATAA	2700
	TAAAGGTCCG CAACCAATTC AACGTATTAG AGACCGCGTA TTAATTGCAA GTAGCAGCAT	2760
35	TTCATATGTT GTAAGTCAAT TAGAGGACAA AGGTTGGATT ACACGTGAAA AGGATAAAGA	2820
	TGATAAACGT GTATATATGG CTTGTTTAAAC TGAAAAAGGT CAAAGTCAA TGGCAGATAT	2880
	TTTcCCTAAG CATGCTGAGA CATTAAACAA AGCGTTTGAT GTGTTAACAA AGGATGAATT	2940
40	AACAATCTTA CAACAAGCGT TTAAGAACT AAGTGCACAA TCTACAGAAG TGTAAGGCGT	3000
	GCACTAAAAA TTTACATTAA AGTATCTCGA TTTCGAGATA AATGCACTAA AAATATAAAG	3060
	AGGGTATATA AAATGATAAA TAATCATGAA TTACTAGGTA TTCACCATGT TACTGCAATG	3120
45	ACAGATGATG CAGAACGTAA TTATAAATTT TTTACAGAAG TACTAGGCAT GCGTTTAGTT	3180
	AAAAAGACAG TCAATCAAGA TGATATTTAT ACGTATCATA CTTTTTTTGC AGATGATGTA	3240
50	GGTTCGGCAG GTACAGACAT GACGTTCTTT GATTTTCCAA ATATTACAAA AGGGCAGGCA	3300
	GGAACAAATT CCATTACAAG ACCGTCTTTT AGAGTGCCTA ACGATGACGC ATTAACATAT	3360
	TATGAACAGC GCTTTGATGA GTTTGGTGTT AAACACGAAG GTATTCAAGA ATTATTTGGT	3420
55		

	TTAAATGAAG GGGTAGCACC TGGTGTACCT TGGGAAGAATG GACCGGTCC AGTAGATAAA	3540
	GCGATTTATG GATTAGGCC CATTGAAATT AAAGTAAGTT ATTTTGACGA CTTTAAAAAT	3600
5	ATTTTAGAGA CTGTTTACGG TATGACAACT ATTGCGCATG AAGATAATGT CGCATTACTT	3660
	GAAGTTGGCG AAGGAGGCAA TGGTGGCCAG GTAATCTTAA TAAAAGATGA TAAAGGGCCa	3720
10	GCaGCACGTC AAGGTTATGG tGAGGTACAT CATGTGTCAT TTCGTGTGAA AGATCATGAT	3780
	GCAATAGAAG CGTGGGCAAC GAAATATAAA GAGGTAGGTA TTAATAACTC AGGCATCGTT	3840
	AATCGTTTCT ATTTTGAAGC ATTATATGCA CGTGTGGGGC ATATTTTAAT AGAAATTTCA	3900
15	ACAGATGGAC CAGGATTTAT GGAAGATGAA CCTTATGAAA CATTAGGCGA AGGGTTATCC	3960
	TTACCACCAT TTTTAGAAAA TAAAAGAGAA TATATTGAAT CGGAAGTTAG ACCTTTTAAT	4020
	ACGAAGCGTC AACATGGTTA ATTGGAATGA GGAGGATTG TGATGGAACA TATTTTAGA	4080
20	GAAGGACAAA ATGGTGCGCC AACACTAATA TTATTGCATG GTACAGGTGG TGATGAGTTC	4140
	GATTTATTAC CGTTAGGCGA AgcATTGAAT GAAAATTATC ACTTGTTAAG TATTAGAGGA	4200
	CAAGTTTCAG AAAATGGGAT GAACCGTTAT TTCAAACGTC TTGGTGAAGG TGTTTATGAT	4260
25	GAAGAAGATT TGGCATTTCG TGGACAAGAA TTGTTGACGT TCATTAAAGA AGCTGCTGaA	4320
	CGTTATGATT TTGaTATTGA AAAAGCAGTA CTTGTTGGAT TTTCAAATGG ATCAAATATA	4380
30	GCGATTAAC TAATGTTGCG TTCAGAAGCA CCATTTAAAA AAGCATTGTT ATATGCACCG	4440
	TTATACCCAG TTGAAGTAAC GTCAACAAAG GATTTATCAG ATGTCAGTGT GTTGCTTTCT	4500
	ATGGGGAAAC ATGATCCAAT TGTGCCATTA GCTGCAAGTG AACAAGTCAT TAACTTGTTT	4560
35	AATACACGTG GGGCACAAGT CGAAGAAGTT TGGGTGAAGG GCCATGAAAT TACAGAACT	4620
	GGATTAACGG CTGGTCAACA AATACTTGGG AAATAACAGT TCTATTAAGA AGCGGACAGA	4680
	TGGAaAAGAT TTTTACTTTT CATCTGCCCG CTTTTTTGAT TTTGAAGTGC TGTAATAAT	4740
40	TTTACAATAG TATAGATATT TTAATCGATA TGAGATTTGC CGGTAATACG CTTAATTAAA	4800
	CCTTTATAGA GTACAGGTAT GAGTAAGATG AAACCGAACA ATCCATAAT AGGGAATACT	4860
	TTTCCAATTA ATGAAATGaA ACCGATAAAT GACTAATAT AAGTGATGAC AGCCATTGTA	4920
45	ATAATAATGA TGAAGTAACG TCTGCTGAAT GGAACGCTGA AACGTGACGC AAATGCATAC	4980
	ATTAATCCAA CAACAGTATT GTAGATGACA AGTATCATAA TGACAGACAT AATAATACCA	5040
50	ATTGACGGAG ACATTTGTGT CGCTAATTTT AATGTAGGTA GATCTACGTG TTTAATTTTA	5100
	TCGAATTGAG AAATTAAACC TAGATTAATC ATCATGAGTA AAAATGTAAT GATTAAACCG	5160
55	CCAATCAAGC CCCCCTATAA CGTTGAGTCA CGATATTTAA CTTTACTACC CATCACTGAT	5220



CCAGGTGATA ATGATTTCTG CTTATGAATC TGAGCATCAT TATTAGCGGC AGTAAATCA 5340  
 AGATGACTTG TTGTGAAATA GTAGACCGCA ATCATAATGA CAATCGCAAT TAAAAATGGG 5400  
 5 GTAACACCGC CAAGCACAGC AATTAAACGA TCGAATTTTA GAAACAGTGT TGCTAAAATA 5460  
 AAGGCGACTA ATATGAGTGC GCTCAGCCAA TACGGTAAGT TGAAACTTTG ATGAATGGTT 5520  
 GACGCACCAC CTGCAGTCAT AATAATAGCT AAAGACAACA TAAACATTGT TAAAAATAATA 5580  
 10 TCAAAACCTC TTGCAATAGA GGGGTATAAG AAATAGTTAA TTGAATCAGA ATGATTTCTG 5640  
 GACTTTAGAT GATGACCTGT ATGCATGACA ACCATTCCAC CTAAAGTAAT CAATAGTCCT 5700  
 15 GTTACAATAA TGCCTGAAAT GCTATATGCG CCATGACTTG TGAAAAACTG GAAAATTTCT 5760  
 TGACCAGTAG CAAAGCCGGC ACCAACGACA ACACCAACAA AGGCAAATGC CACAATAATG 5820  
 GACTCTTTTA AGATACGCAT GATTTAAAAA TGTCCTTCG TAATTTTAAG TAATATAGAA 5880  
 20 AATGTAACAT ACATGTTAAT GAAAAATATA GTACTAATAT AGTATTTTGT TAAATTGGAG 5940  
 TAGAAGCGAG GGTGTCGGTC ATTCATTAA TTTATTAGTT GATTTTGCAT TTTTGTGCTG 6000  
 TAAAGTTGTT ATAATACAGT TAACAGGAAT TAGCATAGAT ACACCAATCC CCTCACTACT 6060  
 25 CGCAATAGTG AGGGGATTTT TTTCGGTGTA GCTAGGTCGC CTATTTATCA TCGTGTGTC 6120  
 GTAGCaATGC GTAAACACAG TACCACTAAA TAAGTGCACG ATACATGCAT CAAATGTCGT 6180  
 CTTTAGTcTA AGTAACGATC ATGCATTAA ATTTTCAAAA TATCTATTTG AGCTTGAAGA 6240  
 30 TCTTTACCAA TATTGGTATC ACGAATCTTC TTACGTTGTA ATTCTTTATC TACGACGCGC 6300  
 TTTATAGAAA GTTCATCGAT ACCTTCGGAA AGTATTTTTn CTTTAGCGTT AAATTGTTGG 6360  
 35 TGTGCAACGA GTTGCAATACC GAATGAATTA TACAATAGTG TATAGCCTGC AATGCCAGTn 6420  
 GTTGACTGAT AAGCTTTTGA AAAGCCACCA TCAATGACAA GCATCTTTCC ATCAGCCTTG 6480  
 AT - 6482

(2) INFORMATION FOR SEQ ID NO: 53:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16592 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

ATTTAAGGCG ATTGCTTG TG TATTTCTCTC TTTTGTAGGC AAACCTGCAC TCGTTCCAAA 60  
 AAATGTAAC TCCATATATG CCCCTCCTTT TCTTCAATTC ATTTTATCAT AAAATTTGTA 120

	AATTTTCTA ACTTTAACGT AGACATAACT ATATAAATTT TGATAATTAC GTTATACTTA	240
	TCATTAATAA GTATCACATT AAACATGATA CATGAATCGA TATTTTCATTT AAGACACTGC	300
5	ATACAGTCGA GCATATTGTA TGACCTACTG AATGGATTAT CTTATAATAA TAAATCATAT	360
	ATCTAATTAA GAATTGAGGT TTTAATCTTG AGTACTAAAA ACAAACACAT CCCATGTTTA	420
10	ATCACAATCT TTGGTGCACT GCGTGACTTA AGCCATCGTA AGTnGTTTCC ATCAATATTC	480
	CATCTCTACC AACAAAGACAA TTTAGATGAA CATATTGCCA TcATCgGTAT TGGACGTCGT	540
	GACATkwnTA ATGATGATTT CCGTAATCAA GTAAAATCAT CAATTCAAAA GCACGTAAAA	600
15	GATACAAACA AAATTGACGC GTTTATGGAA CATGTCTTCT ATCATAGACA TGATGTTAGT	660
	AATGAAGAAA GCTATCAAGA ATTACTAGAT TTTAGTAATG AATTAGATAG CCAATTGAA	720
	TTAAAAGGTA ATCGACTATT CTATTTAGCA ATGGCACCAC AATTCTTTGG CGTTATTCT	780
20	GATTATCTAA AATCTTCTGG TCTTACTGAT ACAAAGGAT TTAAACGCCT TGTTCGAA	840
	AAACCATTCG GTAGTGATTT AAAATCAGCC GAAGCATTAA ACAATCAAAT TCGTAAATCA	900
25	TTTAAAGAAG AAGAAATTTA TCGTATTGAC CACTATTTAG GAAAAGACAT GGTTCAAAA	960
	ATCGAGGTAT TACGTTTTGC GAATGCGATG TTTGAACCAT TATGGAATAA CAAATATATT	1020
	TCAAACATCC AAGTTACATC TTCTGAAATA CTAGGTGTTG AAGATCGTGG TGGTTATTAT	1080
30	GAATCAAGTG GCGCGCTAAA AGATATGGTG CAAAACCACA TGTTACAAAT GGTTCgATTA	1140
	TTAGCTATGG AAGCACCTAT TAGTTTAAAT AGTGAAGATA TCCGTGCTGA GAAAGTAAAA	1200
	GTACTTAAAT CACTGCGTCA TTTCCAATCT GAAGATGTTA AAAAGAACTT TGTTCGTGGT	1260
35	CAATATGGCG AAGGCTATAT CGATGGTAAA CAAGTTAAAG CATACCGTGA TGAAGATCGC	1320
	GTTGCAGATG ACTCTAACAC ACCTACCTTT GTTTCAGGTA AATTACAAT TGATAACTTT	1380
	AGATGGGCTG GTGTACCATT CTATATTCGT ACTGGTAAAC GTATGAAATC TAAAACAATT	1440
40	CAAGTTGTCG TTGAATTTAA AGAAGTACCA ATGAACCTAT ACTATGgAAA CTGaTAAACT	1500
	GTTAGATTCA AACCTATTAG TAATCAATAT CCAACCTAAT GAAGGTGgTA TCTTTtACAT	1560
	CtAAATGcTA AGaAAAATAC ACAAGGTATC gAAACAGrAC CTGtCCmATT GtCTTACTCm	1620
45	ATGaGCGcTC aAGaTAAAAT GaATACTGTA GATGCATATG AAAATCTATT ATTTGATTGT	1680
	CTTAAAGGTG ATGCCACTAA CTTACGCGAC TGGGAAGAAT TAAaATCAAC ATGGAAATTT	1740
50	GTTGATGCAA TTCAAGATGA ATGGAATATG GTTGaTCCAG AATTCCCTAA CTATGAATCA	1800
	GGTACTAATG GTCCATTAGA AAGTGATTTA CTAATTGCTC GTGATGGTAA CCATTGGTGG	1860
55	GGACGATATT CAATAATTGA ATTAAAACGC ACATGTTAAA CAAAAATAAA TGAGCGAATG	1920

	TATATTATGA AATTATATTT TACAATGCCC AAAACTATTT TAATAATCAT TGAACAAATG	2040
	GGTGTATAAT TTATAGAAAT AATGTAGAAT AAAAATAAAT GATTGAATTA ATTGGAGTGA	2100
5	AAGTTTTGGA CGTTATCAAG CAAATACAAC AGGCAATTGT TTATATTGAA GATCGTTTAT	2160
	TAGAGCCTTT CAATTTGCAA GAATTAAGTG ATTACGTTGG TCTTTCGCCA TACCATCTTG	2220
	ATCAATCATT TAAAATGATT GTCGGCTTAT CTCCAGAAGC TTATGCACGC GCGCGTAAAA	2280
10	TGACACTCGC TGCAAATGAT GTGATTAATG GTGCTACACG ACTTGTAGAT ATCGCTAAAA	2340
	AATATCACTA TGCAAATTCA AATGATTTTG CAAATGATTT TAGTGATTTT CACGGCGTAT	2400
	CACCTATTCA AGCCTCTACT AAAAAAGATG AATTACAAAT TCAAGAGCGA TTATATATCA	2460
15	AATTATCAAC TACTGAGAGA GCACCTTATC CATAAGATT AGAAGAGACA GATGATATTT	2520
	CATTGGTTGG ATATGCACGA TTTATAGACA CTAAGTATTT GTCACATCCT TTTAATGTTC	2580
20	CGGATTTTTT AGAAGACTTG CTCATTGATG GTAAAATTAA AGAGTTACGA CGATATAATG	2640
	ACGTTAGTCC ATTTGAACTA TTTGTTATTA GTTGTCTCTT TGAAAATGGT TTAGAAATAT	2700
	TTGTAGGTGT ACCAAGTGAA CGTTATCCTG CACACTTAGA AAGTCGATTT TTACCTGGCA	2760
25	AACATTGTGC GAAATTCAAT TTACAAGGTG AAATTGATTA TGCAACTAAT GAAGCTTGGT	2820
	ACTATATTGA ATCAAGTTTG CAGTTAACAT TGCCATATGA ACGAAATGAT TTATATGTTG	2880
	AAGTGTAACC TCTCGATATT TCATTTAATG ACCCATTCAC TAAAATTCAG CTTTGGATTC	2940
30	CTGTTAAACA GAGTCCTTAT GACGAAGATT AAATAATAAA AAACAAAGAA GCCCCCTAAT	3000
	ATATCTATAG GTCTACAAAT GGCCTTAGAT TCTATTAGGG GGCATATTAA TATGTTAATT	3060
35	TAGTTCGATA ACACATGCTT CATATGGACG TAACTGTTTT AAATTAACCT TGGCATCATA	3120
	ATTAAATAGC TTTACTTCTC CATGGCTTAA ATCAAATGGT ACAGTTAATT CTGCTTCGTG	3180
	GTTAGTAAGA TTACCTACAA TAAGAACTTG CTTTTCATTT AATGTTCTCG TGTACGCAAA	3240
40	AACTTGTGAA TTTTCAGCAT CTAATAATC AAATTGACCA TATACGTATA CATCATAGA	3300
	CTTCTTAAT TGAATTAAAT CTTTATAAAA TTGTAATACT GAATGCTCAT CTTCTAATTG	3360
	TTGTGCAACA TTGATAGTTT TATAATTCGG ATTCACTGGG AACCACGGTT CACCATTGTG	3420
45	AAATCCTCCA TTTAACGTAT CATCCATTG CATTGGTGTG CGAGAATTAT CTCGGTTCTC	3480
	ATCTTTATAT TTCGCAAGTA AAGCGTCTAC ATCTCCACCT TGAGCTTTCA CTATTTGATA	3540
	GTCATTTTTA ACAGCAACAT CGTTAAACGT TTCAATACTT TCAAATGGAT AATTCGTCAT	3600
50	ACCAATTTCT TGACCTTGAT AAATGAATGG CGTACCTTGT TGCAAGAAAT AAACAGCTGC	3660
	ATGACTTGTT GCTGATTCAT ACCAATACTT GTCATCGTCA CCCCACGTCG ATACACGTCG	3720
55		

	CCATCTATTT AATACAGATT TATACGAATT TACATCAAAG TGAGAATCAC CACTATTCCA	3840
	CAGTCCCAAA TGTCAAAT GGAATATCAT ATTAAATTTA CCATTTTCTT CCCCACCCA	3900
5	GTCATCAGCA TCATCAGGGC TTACACCATT CGCTTCACCA ACAGTCATAA TGTCATACTT	3960
	ACTTAATGAG CGATCTTTCA TCTCTTGTA CCAAGTTTGT ATACCTGGCT GATTCATATC	4020
	TACATCAAAT GCTGGGGCAT ATGTTTTACC CTCAGGTACA GGTAAGTCAC CCGCTTCAAA	4080
10	CGTCTTCTTA ATATGCGTAA TTGCATCTAC TCTAAATCCA TCAATGCCTT TATCAAACCA	4140
	CCAGTTCATC ATTTCAAATA CAGCATCTCT AACTTCCGGA TTACCCCAAT TCAAATCAGG	4200
15	TTGTTTTTTA CTGAATAAAT GGAAATAATA TTGCTCAGTA TTAGCATCAT ATTCCCATGT	4260
	AGATCCATTA AATATACTTT CCCAGTTGTT AGGTTCAGAG CCATCTGGCT TTGGATCTTG	4320
	CCAAATGTAC CAATCACGTT TGGGATTGTC TTTACTAGAT TTGGATTCTA TAAACCAAGG	4380
20	ATGTTTCATCA GATGTATGAT TTACAACATA ATCTAAAATA AGCTTCATGC CTCTATCATG	4440
	AACACCTTTT AATAAACGAT CAAAGTCTTC CATCGTTCCA AATTCATCCA TAATCTCTTG	4500
	GTAGTCACTA ATATCATAAC CATTGTCATC ATTAGGTGAT TTAAACATTG GACTGAGCCA	4560
25	AATGACATCG ATACCGAAAT CTTTAAAGTA GTCCAATTTA TCAATCATTC CAGGTAAATC	4620
	CCCAATACCA TCGTGATTAC TATCATTAAT AACTCTTGGA TATACTTGAT ATGCTACTGC	4680
	TTCTTTCCAC CATTGCTTAT TCATTTTAAA ACTCCTTTGC TATCGCTGTG TTGATTTTCT	4740
30	TATTTTAAAT TCTGTATCTA TAATGACGAG TTCAATAACA TCCTGTGCTT TGTTTTTCAA	4800
	TATATTTAAA ATTGCTGCAC CAGCCTGTTG ACCTAACATT CGAGGCTTGA TGTCAATACA	4860
35	GGTTTGGGT GGTGACGCAA TTTCGGTTAA ATAAGAATCA TTGAACGTTG CTGTCATTAC	4920
	ATCTTTCGGA ATTTCAATAT TAAGTTCATA TAGGACACTT AAAATCGCTA AATGTAACAT	4980
	AGCATCTAAC GAAATGATTG CCTGTTTAAAT ATTTGGGTCC TTCAAACGCG TATGTAGATT	5040
40	TTGCATGTAA TTAATAATAA CTCTCTTTT ATTACTAGTC TCAATAATTT GATAATTAAT	5100
	TTTATTTTGA GAAGCTATCG TTTCAAATCC TTGAATTCTA TCTTTTGAAA CTTCAAAATT	5160
	TCCTTTTCT GTAATAAATA TTAATTCATC TACACCTTGT TCAATAACAT GTCGTGTCAA	5220
45	ATTTTCAGAA GCTAATATAT TATCATTATC TATATGTGTA AATTGATGAT CTATATCCGA	5280
	TGTAGGCTTA CCAATCACAA TAAATGGCAT GCTTTCATCA ATTAACATTT GTTTAATCGG	5340
	ATCATTTTCT TTTGAATAGA GCAGTATAAA CGCATCAACC ATTCGTTGTT TAATCATTTT	5400
50	ATAAACTTCA TCCATTAAAT CATTCATATT ATTTGAGACT GTCGTTTGTG TACCATAGCC	5460
	ATGCTGGTTA CACGTTTCAG AAATTCCTAG CAATACATTG ATGTAGAATG GATTGAGTCG	5520
55		

	AGTTCTAGCA GCGGTATTAG GAAAATAATT CAATTCTTCC ATAACTTTCT TCACTTTTGA	5640
	AATTGTGCGCT TCGCTAATAC GTTGATTTCC TTTTATAACT CTTGAAACTG TCGAAGGAGA	5700
5	AACACCGGCT TTTAGTGCAA CATCTTTAAT CGTAACCATT TAATCACCTC CTGTTAATTT	5760
	CTGCATCGGA AAACGCTTCC AACCCTGTA TAATACCAGT TTAGTCACAC TTTCTAAAAA	5820
10	AGTCAAAAGA TTTGTGCAAA CGATTGCATA AAACGATAAA AATAAACCTC TCATACTGAA	5880
	ATTCAAATCCG AAAATCAATA TAAAGGTTTG TATAAATATT AAAATCGATT GTTTAGTCAC	5940
	TAACTGCAAA ATAGTTACCT TGGCCATCTT GAAAATTAAA TACACGTTGA CCATTCAATTT	6000
15	CTACTATATC ATGCCCAGTT AAACCTAAAT CATTTAATTT TGAGTATAAT GCATCAAAGT	6060
	TTTTCTCTTT AAACATTAAA GATGGTGTTC CTAGGTTTCC TTCCGGGCTA TGCTTTTCAA	6120
	TAAATTCCTT TGCCATAATC GTCAATGACG TTTCAGCATC TTTGGTAGGT GATACTTCAA	6180
20	CTGCAACATA GTCCTCAGCT AACGGTGTTC CACTTACAAC AACAAATTCT AAAGTTTCTG	6240
	TCCAAAATGC TTTGCTTTT TCGACATCAT CAACATATAA CATAACTTGA TTTAACTTTT	6300
	CCATAAAATA GTACCTCTAT TTCTCTATAG TACATGCTAT CATAACACAG TAAATATTTT	6360
25	ATTACTTCAC AAAATGCTTA AAAATATGGC GGGATGCTTT TAAGGTCAAG GATAATACTT	6420
	GTGTAATTTT TTATAGGTTG TAGCTACTCT ATCACACTCT CTTTTATATT TATCAAAAGA	6480
	TATAAAAAAG GATAGTATCT TTCAACTATC CTTTAATCAA TATTATTCTT CAATCCATTG	6540
30	TGTATGGAAT ACGCCTCTT TATCTTTTCT TTCGTACGTA TGAGCACCGA AGTAGTCACG	6600
	TTGTGCTTGA ATTAAGTTTG CAGGTAAATC AGCAGCACGG TAACTATCAT AGTAATTAAT	6660
35	ACTTGATGAG AAACCAGGTG TTGGTACACC ATTTTGAACA CCAGTTGCGA CAACATCACG	6720
	TAACGCATCT TGATATTCAG TAACGATGTT TTTAAAGTAA GGATCTAGCA ATAAGTTTGT	6780
	TAATCCTGGA TTATTATCGT AAGCATCTTT GATCTTTTGT AAGAATTGTG CACGGATAAT	6840
40	GCAACCTTCT CTCCAAATCA TAGCTAAATC ACCAAGTTT AAATTCCATT CATTATCTTC	6900
	ACTTGCTTTA CGCATTTGcG CGAAACCTTG TGCATAAGAA CAAATTTTAC TCATATATAA	6960
	TGCTTTACGA ATTTTCTTA AAAAGTCTTT CTGTGCACCA TCAAATGATG CTTTTGGACC	7020
45	ATTTAATTCT TTAGAAGCAT TTACGCGCTC TTCTTTGaTT GAAGAGATAA AACGTGCAAA	7080
	TACAGATTCA GTAATGATTG TTAATGGAAT ACCTAATTCT AATGCGTTAA TTGAAGTCCA	7140
	TTTTCTGTGA CCTTTTGTaC CTGCAGTATC AAGAATTTTT TCAACTAATG CTTCTTTATT	7200
50	TTTCATCTAAT TTCATGAAAA TATCACCAGT GATTTCAAAT AAATAACTTT CTAATTCACC	7260
	AGCATTCAG TCTTTGAACG TTTGAGCAAT GTCTTCATGA GACATGCCTA ATAATTCCTT	7320
55		

	CATTTTCACA TAGTGTCAG CACCAATTAGG TCCAATATAA GTAACACATG AAGCACCGTC	7440
	TTTTGCCTTT GCAGCAATTG CATCAAGAAT ATCTGCAACT TTGTTATAAG CTTCTTCTTG	7500
5	TCCACCCGGC ATTAATGACG GACCAGTTAA CGCTCCAATT TCACCACCAG AAACGCCCAT	7560
	ACCAATAAAG TTGATTGCAC TTTGTGyWAA TGCTTTATTA CGTCTGATAG TATCTTGATA	7620
	GTTTGTATTA CCACCATCAA TTAAAATATC TCCATCATCT AATAAAGGTA ACAAACATATC	7680
10	AATCGTTGCG TCCGTAGCTT TACCTGCTTG AACCATTAAAT AAAATTTTAC GTGGTTTTTC	7740
	TAAAGAATTA ACAAATTCTT CCAATGAATA CGTTGGATGA ATATTTTTCC CTTTGTATTC	7800
15	TTCAACCATT AAATCAGTTT TTCACTTGA GCGGTTAAAT ACAGATACAC TATATCCGCG	7860
	TGATTCAATA TTCCAAGCTA GGTTTTTACC CATAACGGCT AAACCAATAA CTCCAATTTG	7920
	TTGTGTCATA TTAATTACCT CACTTGTTGA TTTTTCATTA GTATTGTATC ACAAATAGA	7980
20	CATACACTAC ACTAAATCAT TTCGAATGTC GCGCAACTAT TTTGATTATT TCTAACCTT	8040
	GACTTGCAAG CAAGTTCAAT GATTTAATCG GCATTCTCTC ATTTGTTGTA TGGATTTTTT	8100
	CATAACCCAC TCCTAAAATG ACTGAAGGAA TACCAAATGT ATTAATAATA CTGCCGTCTG	8160
25	AACCGCCACC AGAAATAATT GTATTTGCAG ATAATCCTAA ATTACGAGCA CTTTCTGTG	8220
	CAATTTTAAC AACCGCTTCA TTATCATTAA TTTTAAATCC TGGATAACTT TGCTCCACTG	8280
	TAACACTGTC TTTCCACCT AATTCTGATG CAGTAGTTTC AAACACATCA GTCATATGTT	8340
30	TGACTGTGT TTTTATTCTT TCTGGATCGT GAGAACGTGC CTCTGCTTCT AAAATGACTT	8400
	CATCTGCAAC AATATTCGTA GCTGAACCGC CATGAACTT ACCAATATTG GCAGTAGTTA	8460
35	TTTCATCAAC TTGTCCTAAT TTCATTGAC TAATTGCTTT CGCCGCAATA TTAATAGCAC	8520
	TAACACCCTC TTTTGGCGTA CTGTCATGAG CGTTTTTGCC AAAAATTTTA GCTGAAATTA	8580
	ACATTTGCGT CGGTGCACCT ACAACCGTAG TACCGACATC AGCACTTGCA TCAATAGCAT	8640
40	AACCAAAGTC CGCGTCCAAC AACTCTGAAT TTAATTCTTT AGCACCAATT AAACCTGATT	8700
	CTTCTCCAAC AGTAATCACA AATTGAATTT GTCCATGTGG GATTTGTTGT TCCTTTATCA	8760
	CTTGCAAAAC TTCAAGCATC GCTGATAATC CTGCTTTATC ATCTGCACCT AGAATAGTCG	8820
45	TACCATCAGA GTATATGTAG CCGTCATCTT TTACAATTGG CTTTACATTA ATTGCGGGTA	8880
	CAACAGTATC CATATGGCTC GTCAAATATA ATTTAGGTAC TTCGCCTTCT TCGATAGTAC	8940
	TATTCATTGT ACACACTAGA TTATTGGCAC CTAATTTAGG ATGTTTAGCC GCTTCATCTT	9000
50	CTTTAACATC TAACCCTAAT GCTATGAATT TTTCTTTTAA AATAGGTTGG ATTGTTGATT	9060
55	CATTCCCTGT CTCAGAATCG ATTTGTACAA GTTCAAAAAA CGTATTAAGT AATCTTTGCT	9120

	GATGAAATAA AATGTTACAG TAATTGACGT TACACAGATT TATCAGGTTT GTAAATTGTG	9240
	TCATATTATT TTCAATTTAT TATATATAAT TATTGTAAGT CAAACTAAGC TTTGTCAAAA	9300
5	ATATATTGAT TGATTTTCA AAGATATCGT ATAATGAGGA AAATGACATA AGCAAACCTTA	9360
	CTCATGTTTT TTATTATATT CCTTTATGAT GATTGCTAGT TATATCGTCT CAAGTTAAAA	9420
	GTTTTATATC TTATGTCGTA ATTATTAATA CAAAGGTTAT TCATTTGGAG GCACACAAAA	9480
10	TGCAAAATAA AGTTTTAAGA ATTATCATT TCGTTATGCT TGTATCAGT GTATTAGCAT	9540
	TGTTATTAAC GAGTATCATT CCAATTTTAT AAATATATC TCAACTACCT ATACAAAATC	9600
15	ATACAATTAA AAATCCATCC ATTATAAAGC CATGTATTAA TAAGTTATCG TATTGCAACG	9660
	ATTACTTTCA AACATGGGTC ATACGGATGG ATTATTTTTT AAGCTACTTC ACTATGCATT	9720
	TTCAATGAAC CAAATTGCGA TTTGATTGT AAATATTCTT CTAATTCATT TAATATTTGA	9780
20	ATAATACTTG CTCTCGAGT AAGCGCTTG TGTGTTGTTG GCAATGGCAG TTCATCCAAT	9840
	TTCAAACGCG TCTCATACAA ATTGTGTAAA CGCATTGCTG TATAGTCATT ACTATTCACA	9900
	TTTAGACCAA TTTCTTTCAG CAGTGACGCA ACATCATTTA AAAGCGGATC TTTATGACAG	9960
25	ATACTTTCGA TGAGCGGTTT CATTCTCATT AACAAATCCA CTTGCTCTTC TCGCATATCA	10020
	AAATAATGAT AGTATGAATT TTCGTTTCTA ACAAATGAT TTTTAACATC TCGGAACGCG	10080
	ATAGACTTCG CCTTTTTAAT ATTTAAAAGT AACACTTCAA ATTCAATCGC AATGGTATCT	10140
30	TCATATTTT CACAAATATA ACTATATTTA CTAAAAATAT CAGCAATTG TTGCTCAATT	10200
	TTACATTTGT ATTCTGCTAG TTGTTGTCT AAACCTGGCA TCATTAAAT CaTTGTAAAT	10260
35	GCAATGCTTA GTCCAATTAA CAGTAATAAT GTTTCATTAA CAATTAAATG TGCATCAATT	10320
	GATTTTGCAT TAAAAACATG AAGTAATATA ACGCAACTCG TAATGACACC TTCTGTACT	10380
	TTTAAATACGA CAGTTAATGG TATAAATAAC AATACGATAA TACCGAGTAC AATTGGACTC	10440
40	TGACCTAATA AACTAAATAT TGCTGAACCT AAAAAACAATA CTAAAAACA TGATACTAAT	10500
	CTTGAAATAA TCGCTGTAG CGAATGTAAT TTTGTATGTT TAATACATAA TACGACTAAT	10560
	ATGGCGCTTG AAGCATAATT ATCTAAACCT AACAGCTTAC TAATAATTAC ACCTAAAGTC	10620
45	ATACCCACTG CTGTTTTTAT TGTCTAAAT CCAATCTTGT AAGGATTTAA CTTTAACATG	10680
	GGTTAGCGCC TCTTATCTTT CTTTACAATA TTTATTGAAT AATGTTTGTA ATTGATTAAT	10740
	TACGTTATC ACATCATGAC CTTGATTTG ATGTCTTTCA ATCATTTCTG TAATCTTTCC	10800
50	ATCTTTTACT AATGCAAATG ACGGACTTGA AGGCGCATAA CCTTCGAAGT ATTACGCGC	10860
	TCTTTGTGTC GCTTCTTTAT CTGTCCAGC AAATACTGTC ACTAGACGAT CAGGTAATAC	10920
55		

	AGAATTGATC ATAAC TAGTG TTGTACCATC TTGTTTAAGA ACTTTGTCAA CATCTTCTGC	11040
	AGTAGTTAAT TGCTCATATC CCGCAGATTC AATTTCAITC CTTGCTTGTT CTACAACACC	11100
5	GTTTCATGTAT AAATCGAAAT TCATGhCCAT AAGTTCAATC ACCTATCCCT TTATATTAA	11160
	ACTATCCTCA TTCTACTAAT TAATAACATA TTGTTCAATA AACTAATCTG AATCACACCT	11220
10	ATATTTAGAC ACAATTTTAA CAATATACCA AACATTATTG TGCTTAAAAT CATGGTAACT	11280
	AATTTGTTCA CATGTTTTCA TTAATATGTT TCAAGTATGA TGTCTTATTT TGACTTTACT	11340
	GCAAAAATGC ATTCAACCAT GTTGATTATT GTTCTTTATC TTTTTTGAAT ATATTGCACA	11400
15	TATTTTAGTG CCAAAAAATA ATACATCCAT CGACAAGAAC AAGATAAAAC AAGTTGTCGA	11460
	TAGATGCATC TATGTTATCA CTAATATATA TTTGTATTTT CTAAAGTATA CTGTTGATA	11520
	CGCTGTTTAA TATGATTCAT AaATTTACCT GTTGTAAAC CATCTAAAAT ACGATGATCA	11580
20	ATTGAAATAC ATAAATTAAC CATGTTACGA ATTGCAATCA TATCATTAAT TACTACTGGC	11640
	TTTTTAACGA TTGATTCTAC TTGTAAAATC GCTGCTTG TGATGATTTAT AATACCCATT	11700
	GATGATACTG AACCAAATGT ACCAGTATTA TTTACCGTAA ATGTACCGCC CTGCATATCT	11760
25	TCAGCTGTCA ATTGCTTATT ACGCGCTTTC GTTGCTAAAG TATTAATTC TCTAGCTATA	11820
	CCTTTGATTG ACTTTTCGTC TGCATGCTTA ATCACAGGTA CGTATAATTT ATTTTCATCA	11880
	GCAACAGCAA TTGAAATATT AATGTCTTTA TGTAAGACAA TTTCATTTCC TTGCCAGCTA	11940
30	CTATTTAATA AAGGATATGC TTTTAAAGCA TCTGCTACAG CTTTTACAAA GAAAGCAAAG	12000
	AACGTTAGAT TATATCCTTC TTTATTTTAA AAGCTGTTTT TATAATGATT TCTCGTATTC	12060
35	ACAAGATTG TAGCATCTAC TTCAATCATC ATCCATGCAT GTGGAATCTC TGTTACACTA	12120
	TTAACCATAT TTTGCGCAAT TGCTTTACGC ACACCATTTA CTGGTATTGT GCTGTTTTCA	12180
	CTATTGCTCT CAGATGATTG GTTACTTGAT GTATCTACTG ATGTTGATTT TGTTTGAAC	12240
40	TGTTTGTGAG ATTGAGCTGT GGTACCACCA TTTTCAATAA CTGACATTAT ATCCTTCTTA	12300
	GTTACACGAC CTTCAAATCC ACTACCTACA ACTTG TGATA AATCAATGTC ATGCTCTGAA	12360
	GCGAGTTTAA ATACAACAGG TGAAAAGCGA CCATTATTAC GTGGTTGATT TTGTTTAGCA	12420
45	GTAGATGTCT GTTCCACTGT TGCACTAGCT TTTT TAGTAG ATTTCTGAGT ATGCTCATCC	12480
	ACTTTTGCTT GTATCTCTTC AGTTGTTTCA TTTGTCTTTT CATCAGCAGT TTCAATTTA	12540
	CAGATAATTG TATCAATAGC TACTGTCTGC CCCGCTTCAA CTAAAATTTT TGTAATTGTT	12600
50	CCTGATATCG TGGAAGGGAC TTCAGCTGTC ACTTTATCTG TAATAACTTC ACATAATGGT	12660
	TCATATTCAT CAATATGATC ACCAACAGAA ACTAACCATT GTTCAATGGT GCCTTCATGA	12720
55		



	AATTCACGCA TTTTATTTAA GATTTTTTCT GGATTCAATCA TAATTTTATT TTCTAATACA	12840
	GGAGAAAATG GCATAGATGG TACAETCTGGA GCAGCTAAAC GCATGATTGG TGCATCTAAA	12900
5	TCGAACAAGC AATGCTCTGC AATAATCGCT GACACTTCTG ACATAATACT ACCTTCTAAA	12960
	TTATCTTCAG TTACAAGTAA AACTTTACCT GTATGTTTAG CACGATCAAT AATTGTTTCT	13020
10	TTATCTAATG GATAAACAGT TCGTAAATCA ACGACTTCAA CATTGATACC GTCTGCAGCT	13080
	AAAATATCCG CTGCTTGTA ACAATAATTG ACCATTAATC CATAACAAAA TACTGTTAAA	13140
	TCTTCACCTT CACGTTTCAC ATCTGCTTTT CCTAAAGGTA CAGTGTAATA TTCTTCTGGC	13200
15	ACTTCTTCCT TTAAGAAACG ATAAGCTTTT TTATGCTCAA AGTACAATAC TGGATCATTT	13260
	GATTCGATAG ATGATAATAA AAGCCCTTTA GCATCATACG GTGTGGAAGG AATAACAATT	13320
	GTTAAACCTG GCGATGAAGC AAATATACTT TCAATACTTT GTGAATGATA TAGTCCTCCG	13380
20	TGAACACCGC CACCAATGG TGCACGAATC GTTAATGGGC ATTGCCAATC ATTATTTGAA	13440
	CGATAACGCA TTTTCGCAGC TTCACTAATA ATTTGATTTG TCGCAGGTAA AATAAAATCT	13500
	GCAAATTGAA TTTCTGCAAT TGGTCTTTTA CCTACCATAG CTGCACCAAT GGCAGTTCCA	13560
25	ACAATATTG ACTCAGCTAA TGGCGTATCG ATAACCTCTGT CTTCAACATA TTTTGTGTC	13620
	AGTCCTTGAG TAGTACCAA TACGCCACCT TTTCTACCAA CATCTTCACC AAGAATAAAC	13680
	ACATCTTTAT TTTGTGTAA TGCTAAGTCT TGTGCCTGCG TATCGCCTCT AAATAAGATA	13740
30	ATTTAGCCAT TAGTTAAGAC TCCCTTCTTC GTACACAAAT GCATAGGCTT CTTCGACACT	13800
	TGGATATGGC GCGTCTTCAG CAGCCTTTGT CGCTTTATTG ATGATGTCTT ThATgTCCGC	13860
35	TTCTATTCT GCCAACCAAG CATCATCGAT AATGCCAGCT GAAAGCAACT CTTTTTTGAA	13920
	CTTTTCATTG CAGTCTGCTT TTTAAGCGT TTCACGCTCT TCTTTCGTAC GATATTGGTC	13980
	GTCATCATCT GATGAATGAG CTGTCATACG ACTTGTTACT GCTTCAATCA AAGTTGAACC	14040
40	TTGACCAGAA ATAGCTCGAT CTCTTGCTTC TTTTCATCGCT TTATACATTG CTAATGGATC	14100
	ATTACCATCT ACTTGTTTAC CATGTATACC GTAACCAAGT GCTCTATCCG ATAATTTTTT	14160
	AGCTGCGTAT TGTAATGAAT CAGGTACTGA AATTGCATAT TTATTATTA TAATGACACA	14220
45	TACAAAAGGA AGTTGTGTA CACCCGCGAA GTTTAAACCT TCATGGAAGT CACCTTGGTT	14280
	TGAGCTACCT TCACCAACAG TTGCTGTTGC AATTTTCTTC TTACCATCCA TTTTAAAGC	14340
	TAAAGCAGCA CCAACAGCAT GGGGTATTTG AGTTGCTACC GGTGAACTTT GAGACAAAAT	14400
50	ATTCTTAGCT CTAATACTAA AGTGTGATGG CATTTGTTTT CCACCAGAGT TAACATCGTC	14460
	TTTCTTTCCA AACGCTGATA AAAACGTATC ATACGCTGAG ATACCCATAT AAGTAACGAA	14520
55		

	AATCTGAGTT GCTTCTTGTC CTTGACCACT TACAACAAAT GGAATTTTAC CTGCACGGTT	14640
	CAATAACCAC AGTCTTTCAT CTATTTTTCT ACCTAAATCC ATCCATTTAT ATATTACTTT	14700
5	TAGGTCTTCT TCGCTAAGGC CTAATGATTT ATAATCAATC ATGTTAAATC CTCCTATTTA	14760
	TACGTGAATA GCTCTACTTT CTGCTTTCAA TCCTAATTCC ATCAACACTT CAGAGATGGA	14820
10	AGGATGTGCG TGTGTTGTTA GTCCTAATTC TAATGCCGAG CCATTCATGA ACTGTAACAG	14880
	TGATGCCTCA TTAATCAATT CTGTTACATG TGGACCAATC ATATTAATAC CCACAATTTT	14940
	TTCAGTTGAT TGATCAATCA CCATTTGCT ATACCCTCG TTTGTGTCAT GGCTATCAAT	15000
15	CACTGCTTTA CCAATTGCTT TAAATGGTAC TTTAAACTT TTAACTTTCA TTCCCTCTGC	15060
	CTTTGCTTGT TCAATGTTTA AACCGATAGA AGCAATTTCA GGTGTGAAT AAATACACTT	15120
	AGGCATCATG TTATAGTTTA CTGGGATTGG GTTCCCCTCA AACATATGAT CAACAGCCAC	15180
20	AACACCTTCT TTTGATCCAA CATGTGCCAA TTGTAATTTT CCTATACAAT CACCAGCTGC	15240
	ATAAATATGT TTATCTTCAG TTTGTTGAAA TTCGTTGTT AAAATATGTC CTGATGTTGa	15300
	AAGTTTATT TTAGTGTTGT TTAAACCAAT ATCTGATGTG TTAGGTTTTC TACCAATCGA	15360
25	TAGCAACACT TTATCTACTT TAATTATGTC TGAGGAAATT TCAAACGTAA CACCATCTTC	15420
	GTTAACATTT ATATCATTTT CAGAAAGTTT TATTCCTCA TAGAATTTAA CACCACGTGC	15480
	TGACAATGAT TTTTTAATA GTTGTAAGC TTGTTTACTT TCAGTTGGTA AAATTCTTTC	15540
30	ACCTGCTTCT ATAACTGTTA CGTCAACACC TAAATCTATC ATCAATGATG CAAATTCCAT	15600
	TCCGATAACA CCACCACCAA TAATACCAAT ACTTGATGGT AACGTCTTTA ATGATAATAT	15660
35	ATCATCGCTA GATAAAATTT TATCATGATC AAATGATAAG AATGGCAACT CTGCAGGCGA	15720
	AGAACCAGTT GCAATTAATA CAAATTGGTT GGGTAATAAG TCTGATTCAC CATCTTCATA	15780
	TTGACAGAA ATTGTGCCAC TTTGAGGTGA AAATATAGAT GTACCTAGAA TACGTCCCGT	15840
40	GCCATTATAA ATGTCAATGT GATTGTGTTG CATTAAATGC TTTACACCTT GATACATTTG	15900
	ATTAATAATG TCTTCTTTTC GTGCCAACAT ATTTTCAAAA TTAACATTAG CATCTTTGAC	15960
	ATCAACGCCA AACATTGCTG CCTGTTTTAC TGTTTGAAAT ACTTCAGCAG ATTTAAGCAG	16020
45	CGATTTAGTA GGAATACAAC CTTTATGGAG ACAAGTACCT CCTAATAGTT GTCGTTCTAC	16080
	TATTGCCACT TTTTACCTA ATTGAGACGC ACGTATCGCA GCAACATATC CTGCAGTACC	16140
	TCCACCGAGA ACGACTAAAT CATATTGTTT CTCTGACATG TTCTTACTCC TAACTAATGA	16200
50	TATATATCCA TTGAAAATTT ATTAATACAT AGTTTTCATG TCCATTAATT ACCTATTTTA	16260
	CATGATTGTC TATTTAGTTT GAATGCACAT AAATAAATCC ATAAATGAGT ATTCAACACA	16320
55		

TAAATCAGTA ACACTTGCAC CTGAAATCAT TCGTGCAATT TCATCTACTT TATCATCGCT 16440  
 AATTAECTCT TGAACCTGTG TTGTTGTACG ATCATCTTTT GATGATTTTG AAATTAATAA 16500  
 5 ATGATGGTCG CTCATCGATG CAACTTGTGG TAAGTGAGAG ATACAAATAA CTTGTATATA 16560  
 TTCTGCTaTA TCTCGCATT TCTCTGCCAT TT 16592

(2) INFORMATION FOR SEQ ID NO: 54:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 13794 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

20 CCAATACAAC GTAAAAAGAT TGCTTGTGTT ATTAATGAGT TAGATAAAAT AATTAAAGGA 60  
 TTTAATAAGG AAAGAGACTA CATAAAATAT CAATGGGCTC CAAAATATAG CAAAGAnTTT 120  
 TTTATACTTT TTATGAACAT TATGTACTCA AAAGATTTTT TAAAATATCG ATTAAATTTA 180  
 25 ACATTTCCTG ATTTATCTAT CTTATATGTA ATATCATCTC GAAAAAATGA GATACTAAAT 240  
 TTAAAAGATT TGTGTGAAAG TATTAGATTT ATGTATCCTC AAATTGTTAG GTCAGTTAAT 300  
 AGATTAAATA ATAAAGGTAT GCTAATCAAA GAACGATCCC TTGCAGATGA AAGGATTGTG 360  
 30 TTAATCAAAA TAAATAAAAT ACAATATAAC ACTATTAAAA GCATATTCAC AGATACTTCC 420  
 AAGATTCTCA AACCAAGAAA ATTTTTCTTT TAAATTTAAA CAGATTTACC TCTTGATAAA 480  
 35 ATAAATAAGC AATCATACTA CTTCTCAATT TAGTATAAAT AAAAATACAT AATTAACTTT 540  
 CTTTTGTTTT TATATTATTT CAATACCCTA CTATATATCA CAACACATAA ATTAAGCATG 600  
 ACAGTCATTC AATTTAGTTC ACCATTTTCG GTTCCAATTT TACTGAGTAT CATGCTTTTA 660  
 40 ATGTTATAAA CCTAATGCTT TAATAAATCG TGTTAATTCT TCTCGCATAC TGTCATCTTT 720  
 CAATGCATAT TCTATGGTAG TTTTAACGAA GCCTAATTTT TCTCCAACGT CATAACGTTC 780  
 GCCTTCGAAG TCATATGCAT ACACTTGGTT ATCATTATTC ATACGTTCAA TCGCATCTGT 840  
 45 TAACTGAATT TCGTTACCTG CGCCTTCTTT TTGCGTTTTT AAATAATCGA AAATTTCAGG 900  
 CGTTAATACA TAACGTCCCA TAATAGCTAG GTTTGATGGT GCCGTACCTT GTGCTGGCTT 960  
 TTCAACAAAC TTTTTCACCT CATACTGACG TCCGTTTTTA GTTAATGGGT CAATAATTCC 1020  
 50 ATAACGATGA GTATCTGCTT CCGGAACTTC TTGGACACCT ATAACGAGT GCCCTGTTTC 1080  
 TTCATAAACG TCAATCAACT GTTTCACCTG TGGCACTTCA GATTCAACAA TATCGTCACC 1140

	TAAACCTTTT TGTTCCTTCT GCCTTACATA AAAAATATTC GCAAGTTCGG TTGAATACTG	1260
	AACCTTCTCT AGTAATTCAG ATTTACCTTT TTCTTTTAAC ACCATTCTTA ATTCTTTTTG	1320
5	ACTATCAAAA TGATCTTCAA TCGCGCGTTT GTGGCGACCT GTCACTATAA TAATATCTTC	1380
	AATTCAGCT CTTGCAGCTT CTTCAACGAT ATATTGTATT GTGGGTTTAT CTAAGATAGG	1440
	AAGCATTTC TTTGGCATCG CTTTAGTTGC TGGTAAAAAT CTAGTCCCTA AACCAGCAGC	1500
10	GGGAATGATT GCCTTTTTTA TTTTTCCTAA AGTTAATGTG CTCCTTTTCC TAAGTATTAA	1560
	ATCTATGTAT CAACGTCATT TTAACACTAA TTAGAACGCC TTCATAGTGT CATTGAGTAT	1620
15	GTAATTATTT CTTGGGAAAT TTGTTTTAAT TTTAAAAAAC AGGCTTACTT CATATAATTT	1680
	ATGAAATAAA CCTGTCAATT TTGGATTGAT TATGCTTTGT GATTCTTTTT ATTTCTGCGT	1740
	AATAACGCTA AACCTAAAT GCTAAATAAT CCGCCGAACA ACATGCCGTT GTTTGTTGAT	1800
20	TCTTCTCCAC CTGTTTCAGG TAGTTCAGAT TTCTTAGATT GTGCTTTTTT AGTTGGTACC	1860
	ACTGCTTTAA CCTTTTCATT GATTTCAATA ACAGGTGTTA CTACTTTACC TTGTTCCACT	1920
	GGTTTAGAAG GTTTTTTAGG TTCTTCTTTA GCAGGTGGTA TTGGTTTACC AGGTTTCAGT	1980
25	GGTACCTCTG GCGTTGGCGG TGTGGTGTG TCCGGCTCGC TTGGTACTTC TGGTGTCGGT	2040
	GGTGTGGTG TTTCCGGCTC GCTTGGTACT TCTGGTGTG GTGGCGTTGG TGGCAGGATT	2100
	GGAGGTGTTG TATCTTCTTC AATCGTTTGT TGACCTTCAT TATGACCACT TACTTGTGGA	2160
30	AGTGTATCTT CTTCAAAGTC AACACTATTG TGTCCACCGA ATTGATAATT TGGTTTATCT	2220
	TTATTTGTAT CTTCTTCAAT AATTTTCAGT TGCTTATTGA ATCCGTGAAT ATGTGGCACA	2280
35	CTGTGGAAGT CGATATCAAT GATATTACCA CCTTGTTCAT ACTTAGGTTT GTCTTTCTCT	2340
	GTATCTTCTT CGAATGATTG GTTACCATTA TTTTGACCAT GAATTTGAGG TACACTATCG	2400
	AAATCGATAT CTACGATATT GCCACCTTGT TCATATTTCTG GTTTATCTTC TTCTGTGTCT	2460
40	TCCTCAAATG ACTGATTACC GCTATTTTGG CCACCTTCGT AACCTAATTC ACTCTTAATA	2520
	TCCACGTGGC TATTTTCTTC GATTTCTTCA ATCAGCCAT AATTACCGTG ACCATTTTCA	2580
	GTTCCTAAAC CAGAATGAGA AATATGATGA TTGTTTTCAG TAATTTCTTC GATTGGTCCT	2640
45	TGCGCTTGAC CATGTTCTTC AGGTAGTTCA TCTACTAGTT CAATCAGATT ACTTTCAGTC	2700
	GTATATTCTT TCGTATCTTC AATTGTTGTA TGATCGCTAA CAGCACCAGT TACAATACCT	2760
	TTTGTAGAAT CTTCTGCAAA TTCAACTAGG TTAGACTCAG TAGTAACCTG ACCACCACCT	2820
50	GGGTTTGTAT CTTCTTCATA TTCAACAACA TCAGCATGAT GTTTTGAATT TTCATGTGTC	2880
	GATTCTTCAA AGTCTACATG AATAGAATCT TCTTCAGTTT CAATGGTACC TTCTGCATGA	2940

55

	TCTTCGATTG TACCACTCAA TTCATGCTTC TCCACTGGCG GCTCTGATTT AAATTCAAGT	3050
5	TCGATAGGAG TACTATGTTT TATAATAGGT TCCTTTAGTT TATCTTTGCC GTCGCCTTGA	3120
	GCGTTATTAG AGTAAAATGC AACGCCATT TCCCaAGTTA AATTACTTGT ATAATAATAG	3130
	TTATAATATC CAAAAAGGTG TGTTCGAAAT TCTAAGTTGC TAGCATTGTA ATCATAATAC	3240
10	CCTTCATATT TTATTACATA ATTTTACTT TGGTCTAAAT TATTAAAGTT TAAAGAATAA	3300
	CCACCATTAG TATCAAAATC TAAACTCATA TTATCAGTCA CATCTTCAAA TTGCTGACA	3360
	TCATCAAGCT TTGCATAnTn AgctTTCAGC TAAATCGTCT GAACCAATGT GTTTATATAC	3420
15	CTTAACTGTT GGATTATTAA CCCCTGGTTT ATTTCTTTA GTTACTTGAC CAGTTACTGT	3480
	CACAGAGCTT AACGACTGGT TGTTAGGTTT CATGTACGCA AAATGACTAA ATTTCCCATC	3540
	TACTTTATTT AAAGTATCAA TTCGACCATT AGCTGTTACT CCCCAATTAT CTCTAACTCC	3600
20	ACCTAAATAT TGAATATTAA ATATTTTGCT AACCGTAGTC TCACCCAATT TAACTTCAAC	3660
	ATTTTGTTA CTTTTTGCG TCACTGTTGT AGGATCAATA AATAGATTTA AAGATAATTC	3720
25	AGCAGTTAAA TCTTCTTTT CTGTACATA TTCTTTAAAC GTATATCTAA CTTTTCTTTC	3780
	TCCAATTATT TCTCCTGTCG CCATAACTTG ACCATCTGTA CTTTTTATCT CCGGAACTTT	3840
	ACGCAGTGTT GAGATACCAT GAGTTTCAAC ATTATCGCTT AATGTGAAAT CAAAATAATC	3900
30	TCCCGCCTTA ATTCCTTCTC CAAATTTCCA TTTATATTTT AAGTTACTC TTTCTGCGTT	3960
	ATGAGGATTT ACAACATTCG TATCTTGTTT ATGTCCTACA ATTTCACTAC CTTCTTCTAC	4020
	TTCCACTTTA TTTGTTACAT CTGTACCTGT CGCTTTAGTT TCTTCCACTA CTTCTTCTC	4080
35	TGCAACTGCT GTAACGTCAt TGatCTTTTC ATTCTTGGTT TAATTTCTGA GACGTTACTT	4140
	GGTTGAGCTA TGTCAACTTG AGTTCCTGTA GTTTCCTTAT CAGCAACTTT TTCCGATGGC	4200
	AAATCAACTC GCGAAgTTTC TACTTTTGCT GCTTGACAG TTTTCGGTGC TTCTTCTGTT	4260
40	GTTACTTGTTG TTGATTGTGA TGGTTGCTCA GTTGATGTCG CGCTGTATGA TTGTGTTTCA	4320
	TCTATTGTAT TAACGTTATT TGTAAGTTGTT TGTGTTTCGC TTGCTTTACT TTCAGTAGCT	4380
	GAACTCCAC TTTCTCTAC TGTAATATTG TTTGTTCCG ATGCTGCAGC TTCTTTTCT	4440
45	TGTCCCATTC CAACAACGAT CATTGTTCTT AAGAATACTG AGGCCGCTCC CAATTTGTGT	4500
	TTTCTTATGC CGTATCTAAG ATTGCTTTTC ACTATAATAT TCTCCCTTAA ATGCAAAATT	4560
50	CATTTATTTT TAAAACTCAA TAAATGCAAT TCTATATTGT TCGGTTTTTA AAAGCAATGA	4620
	AAAAAGCGA GTTAATAAAA AGTTAAGATT GTTGTTAACT TTATGTATAA TGAGTTTTTT	4680
55	ATTATTTGAA ACTCACATAT ATATTGCATA CAAAGCTCTT GAACACCTTG ATATAACAGG	4740

	TACTAAACCA TACATAATAA TCGCCTGTAC AATGCATCAT TAACAAGTCA CTGAAACGCC	4860
	TTTCATTGTA TTAATAACGT CACTATAATT TTTATATCGT TCGGTTTTTG TTTGATTTTA	4920
5	ATGATTATTT ATACAAAAAC AGCCGTATTT CAAGCCGACA TTTTAAATTT AACTAAATTT	4980
	GCATCTAGTT AATAATTGCA TTTATCAAAT TTGTCTTATT GATCCAATCT AATTTGTACT	5040
	CACAAACTAG TTTAAAATTC TAACCTTATC TCTCAGTTCT TTATCAATCA TCAGACATAA	5100
10	ACCAATGAAG CAATCAGAAA AACTCTAAT TTTCTATTAG AAATTTGATT TAATATAAAA	5160
	AAACAGGCTT ACTTCATATA ATTTATGAAA TAAACCCGTC AATTTTTGTT TAATTATGCT	5220
15	TTGTGATTCT TTTTATTTCT GCGTAATAAT GCTAAACCTA GAATGCTGAA TAATCCGCCG	5280
	AACAACATAC CTTTGTGTTG TGATTCTTCT CCACCTGTTT CAGGTAGTTC AGATTTCTTA	5340
	GATTGTGGTT TTTTAGTTGG TGCCACTGCT TTAACCTTTT CATTGATTTT AATAACAGGT	5400
20	GTTACTACTT TACCTTGTTT CACTGGTTTA GAAGGCTTTT TAGGTTCTTC TTTGGCAGGT	5460
	GGTACTGGTT TACCAGGTTT AGCTGGTACC TCTGGTGTG GCGGTGTTGG AGTTTCTGGC	5520
	TCACTCGGCA CTTCTGGTGT CGGTGGTGTG GGTGTTTCCG GCTCACTTGG TACTTCTGGT	5580
25	GTTGGTGGCG TTGGTGTTC CGGCTCACTT GGTACTTCTG GTGTCGGTGG CGTTGGTGGC	5640
	ACGATTGGAG GTGTTGTATC TTCTCAATC GTTGTGTGAC CTTCATTTTG GCCGCTTACT	5700
	TTTGAAGTG TATCTTCTTC AAAGTCAACA CTATTGTGTC CACCGAATTG ATAACTTGGT	5760
30	TTATCTTTAT TTGTATCTTC TTCAATAATT TCAGTGTGCT TATTGAATCC GTGAATATGT	5820
	GGCACACTGT CGAAGTCGAT ATCAATGATG TTACCGCCAT GTTCATACTT AGGTTTGTCT	5880
35	TTTTCTGTAT CTTCTCGAA TGA CTGATTA CCTTTATTTT GACCATGAAT TTGAGGTACA	5940
	CTATCAAAAT CGATATCTAC GATATTGCCA CCTGTTCAT ATTTAGGTTT GTCTTCTTCT	6000
	GTGTCTTCCT CGAATGACTG GTTACCGCTA TTTTGGCCAC CTTCATAACC TAATTCACTC	6060
40	TTAATATCAA CGTGGCTATT TTCTTCGATT TCTTCAATCA CGTCATAATT CCCGTGACCA	6120
	TTTTCAGTTC CTAAACCAGA ATGAGAAATA TGATGATTGT TTTTAGTAAT TTCCTCGACT	6180
	GGTCCTTGTG CTTGACCATG CTCTTCAGGT AATTCATCCA CTAATTCAAT CAGATTACTT	6240
45	TCAGTTGTAT ATTCTTTCGT ATCTTCAACT GTTGTATGAT CGCTCACCTGC GCCAGTTACA	6300
	ATACCTTTTG TAGACTCTTC GTCAAATTCA ACTAAGTTAG ACTCAGTAGT AACCTGACCA	6360
	CCACCTGGGT TTGTATCTTC TTCATATTCA ACAACATCAG CGTGATGTTT TGAATTTTCA	6420
50	TGTGTAGATT CTTCAAAGTC AATTGGATTT GATTCCTCAG AGGACTCAGT GTATCCTCCA	6480
	ACGTGACCTG CTCGCTATC CACAGCAGTA TGGTAATCGA TATCAATAGC TGATGAATCC	6540

55

	TGGTAATCAA TGTCAAGAGT TGATGAATCA TATTCCTCTT CAACAGTAGT TACTAAATTC	6660
	TTATCATATT GACCTGTAAG AGTTTCTTTA ATTGTATCTT CTTTATATTC AAATTTATTA	6720
5	TTTTGAATAA TCGGACCATT TTTCTCAFTT CCGTTCGCTT TATTACTGTA TAAAACTAAA	6730
	CCATTATCCC AAGTTAAGGT ATATCCTCTA TCATAATAAT ACTTATAAAG TTGCTCTGGA	6840
10	TGTCCTACCA TTTGTGTTCT AAAATCAACT TCATCAGTAC CATTAAATA CTCTCCATCA	6900
	TAGTGAACAA CATAAGTTTT ATCTAGATTT TCTATATTCA ATGAATAGCT TCCATTATTT	6960
	TGTAAATTCA AATTCCTACT CATATTACTT GTGACTTCTT TAAATTTAGA AGTATCTGTC	7020
15	GTATTGCAT ATACACTCTT CGCTATGTCT TCATTATTAC CCAAGTATTC AAATATCCTA	7080
	ACTTTTGGTT GATTTCCATT CTGATTACTA CCTTTCATTA AAGTTCAGT AACAGTCACA	7140
	CTTGTCGTTT TACCATTATT AGGTTTAATA AATGCAACAT GCGAAAATCT ATTATTCGCT	7200
20	TTATTAAATG TCTCAATCGA TCCATTTAAA TTGGCATAAT AATTCCTAAT ACCATCTTTA	7260
	TATTTAACAT CTAATTCCTT TGAAGTTTGT TCTTCATTTA GTGTTGAAGT TATAGTTTGA	7320
	TTTCCATTAG TTTGTACAGT TTTAGGATCA ATAAATAAAT TAATTTCTAG TTCAGCCGTT	7380
25	ACATCAACCT TATCTTCAAT ATCATTTGTA AATGTATATC TAATCTTTCC ACCTTCTAAA	7440
	ACTTCACCTG TCGCCATTAC GACTGAACCA TTTTAAATTT CTGGTACTTT TCTAGCAGTT	7500
	GATACGCCAT GCGTATTTAC ATTATTTGAT AAAGTAAAGT CAAAGTAGTC ACCTTGATGT	7560
30	AAACCATTCT CAAATTTCAA CTTATATTTT AGTACCGCTC GTTGTCTGTC ATGAGGTTCT	7620
	ACTTTATTTG TATTGTTATG CCCCTCAATA GAACCAATTT CTAAGTAAAC TTTACTTGTT	7680
35	ACATCTGTAC CCGTTTCCAC TTTGCGGTTA CTAGCTTCCT TAGCTTCCGC TACATCTGCT	7740
	GATCTTGTC CACGTGGCTT ACTTTCTGAT GCCGTTCTTG GCTGTGCCAC TTCAACTTGT	7800
	GTTTCTGCGA CTTGATTTTG TGTAGCCTTT TTAGGTGTTA AATCTACTTG TCTTTGATCT	7860
40	CCGCTATTGT CTTGAGATTG TGTTGTTTCC TTAAGTTGAG GTTTCGCTTC TTCCTTAACT	7920
	ACCTCTTCTT TAACTGTTTC TATATTTGCT GGTTGTGCAG TTTGTGGTGC TTGTACTGCT	7980
	TTTGGTGCTT CTTCAATTGT TACTTGTGTT GCGTTTGACG GTTGTCTGT TACTGTTGCG	8040
45	TTATATGATT GAGTTTCTTC TATATGATTA ACGTTAGTTG CAGTTGTTTG TGTTTCACTT	8100
	GTTTATTTAT CAGTAGCTGA ATTCCCATTT TCTTCTACTG TAGTTGTCTT TTGTTCTGAT	8160
	GCTGCAGCTT CTTGTCTTG TCCCATCCCA ACAACGATCA TTGTTCTTAA GAATACTGAT	8220
50	GCTGCTCCCA ATTTATGTTT TCTAATGCCG TACCTAAGAT TGTTTTCAC TATAATATCT	8280
	CCCTTTAAAT GCAAAATTCA TTAATTTTTT AAACCTAATA AATGCAAGTC TATATTGTTT	8340
55		

	ATGTTAATTG ATAATTTTAT TATTTGAAAT ATACCTATAA ATTGTATTCA AGTCATCAGA	8460
	AACCCCTGTC ACACAAGGCT TGTATTTTTT ATACTTATTT TTAAATTAA ATTCATCATT	8520
5	ATCTAATTTA AAACAATATA CTAAACGTTT CATAATTATC GCCTGTACAA TACGCACAAA	8580
	AACATGTCTT GAAACGCCTT TCATTACTCT AAAATACCCA ATATACTTTT TATATCGTTC	8640
	GGATTCTGAG TATTTGAGAC GATTTTCTGC ATAAAAATAA ACGTGTTC AAGCAATATA	8700
10	TTGCAATTAC CTAAAAACAC GTTTACTTAA TATTTAGTTA AACAAATAAG CTAATGAATA	8760
	AAATGAAGAT GATACCTGAA ACGGAAATAA TCGTTTCTAA TAATGACCAT GTTAAGAATG	8820
15	TTTCTTTTAC AGTTAAACCA AAATATTCTT TAAACATCCA AAATCCTGCG TCATTTACAT	8880
	GAGACAAAAT CACACTACCT GCACCTATCG CAAGTACAAC TAATGCAACA TTTACATCTG	8940
	ATGATTGTAA TAATGGTAAG ACAATACCTG TAGTTGAAAT CGCAGCTACT GTAGCCGAAC	9000
20	CTAATGCGAT ACGTAGCACA GCTGCAACAA TCCATGCTAG TAAATCGGA GACATCTCTG	9060
	TACCTTCAAA CATTTTAGCA ATTGTATTTT CGACACCGCC GTCAATTAAT ACTTGTTTAA	9120
	ATGTACCGCC ACCGCCAATA ATCAATAACA TCATTCCGAT TGGATAAATC GCATTGTC	9180
25	CTGATTCCAT AATATGATTC ATCTTACGCT TTCTCATTAA TCCCATCGTA ACGATTGCAA	9240
	ATAATACTGC TATTAGCATG GCTGTCCCTG CTGTTCTTAT CATATAAATG ATAGATTCAA	9300
	ATAGATTTGT AGGTTTGTCA TGCCAGTTA CAAGTTGCGT TATCGTAGAC ACTAACATTA	9360
30	ATATGACTGG TAATGTTGCT GTTAATAAAC TCATACCAA TCCTGGCATC TCTTGATCCG	9420
	TAAATCTTTT TTGTGCACCT AACGCTGAAA TATCGCCTTC TCGTGTATAC GCAGACGGAA	9480
35	TCATTTTTTG TGCACTTTGT TAAATATAGG CCCTGCAATG AGTGTAATG GaATGGCAAT	9540
	AATCATACCA TACAGTAATA CATCTCCAAC ATTTGCCTTT AATTCTTTTG CGATGACTAC	9600
	CGGTCCTGGA TGTGGTGGTA AAAAGCCATG TGTCATGAT AAAGCTGTTA CCATAGGTAG	9660
40	TCCTAGTTTT AACACTGAAA CATTTGCGCG TTTTGCTACT GTAAATACTA ATGGAATCAG	9720
	TAAGACTAAA CCTACTTCAA AGAACAATGC AATACCGACG ATAAATGCTG CAACAAGCAT	9780
	TGCCCATTGT ACATGTTTTT GACCAAATTT TTGAATCAAC GTGTCTGCGA TTCGAGTTGC	9840
45	ACCACCACCA TCAGCAAGCA ATTTCCCAAG TATGGCACCT AAACCGAATA TCAGTGCAAT	9900
	GTGGCCGAGC GTACTGCCCA TTCCTTTCTC AATCGTCTCC ATAATTTTAG TCAATGGTAT	9960
	ACCTAGCATT AACGCTGTAA TCATCGATGT GATAATTAAT GAAATAAATG TATTTAATTT	10020
50	AAACCCAATA ATTAATACTA ATAAAATAAC GATACCTAAA ACAACACTGA TTAACGGCCA	10080
	TATTTGTTA AACATGACAT TCCCCTCTTT CTCTTTTCAA TAGAATGTAA CACCGTCGTC	10140
55		



	GAGTGACGTA TTTATTGTGT TTTATTTTCA GCGATATGTT GGC GTTGAAA ATCTGCAATT	10260
	TGTT CATAAT TCTCTGTTAA AGAACGACTT AAATTGATAA AAATGGATAC GATCTCTTGG	10320
5	TAAACAGTGA CATTTTCTTC AATCGGCCGTG TGATTGTTTG TGGCACCGAC CATCGATGAA	10380
	ACGATTGAAA AATCTTCAAT GTCACCTACA GCTTTAAGTC CGAGCACGCA GGCACCTAAG	10440
10	CATGAAC TTT CATAACTTTC AGGAACCACT AACTCTGTGT CAAATATATC TGACATCATT	10500
	TGACGCCATA CTTCACTTTT CGCAAAACCA CCTGTTGCTT TTATCATCTT AGGTGTTTCA	10560
	TTCATTACIT CAATAAGCGC AAGATAGACG GTATACAAAT TGTAAGAAGAC ACCTTCTAAT	10620
15	GCAGCGCGAA TCATATGTTT TTTTITATGA GATAAAGTTA AACC GAAGAA TGAACCTCTT	10680
	GCATTTGCGT TCCAAAGCGG CGCACGTTCT CCTGCTAAAT AGGGATGGAA TATTAAACCA	10740
	TCTGCACCTG GTTTAACACG CTTTGCAATT TGAGTTAAGA CATCATAAGG ATCAACACCG	10800
20	AGACGTTTCG CAGTTTCGAC TTCACTCGCT AGCAACTCGT CGCGCAACCA TCTCAATACG	10860
	ACACCACCAT TATTTACAGG ACCTCCGATG ACGTAGTGGT CCTCTGTTAA GACATAACAA	10920
	AATATTCTAC CTTTGTAATC AGTACGCGGT TTATCTATCA CAGTACGAAT CGCCCCAGAT	10980
25	GTACCGATTG TGACAGCAAC TTCTCCTTTA CCAACACTAT TGACACCTAA ATTAGAAAGG	11040
	ACCCCATCAC TCGCACCAAT AACAAACGGT GTATCTTTAT TAAGCCCCAT TAATGTTGCA	11100
	TAACGTTCTT TCATACCTTT CATCACATAC GTTGTGGA CTAATTCCGG CAACATTTCC	11160
30	TTGGAATAAC CCAGCAGTTC TAATGCCTCA ACATCCCAAT CTAATGTTTC TAAATTAAAC	11220
	ATCCCTGTTG CGGAAGCCAT TGAATAATCA ATGATATATG TATCAAATAA ATGATAGAAA	11280
35	ATGTATGTTT TAATATCTGC AAACCTTAGCA GTACGTTGAA ATACATCTTG CCATTCATGT	11340
	TTCATCCAAA AAATCTTCGC TAATGGCGAC ATAGGATGAA TCGGTGTGCC TGTTCGCTGG	11400
	TAAATCGCAT TGCCATCATG CACTTCATTT ATTACTGTTG CATATTTTGC AGCGCGGTTA	11460
40	TCTGCCCAAG TAATATTATT TGTTAATCTT TGATGTTGCT GATCCATCGC AATCAAGCTA	11520
	TGCATTTGCG CACTAAATGA CACAACTTA ATGTCGTCTT TATTAAC TTT GGATTCTCTC	11580
	ATAACATATT TAATAGTCAT TAGTACTGCA TCAAATAATT CATCTGGGTT TTCTTCTGAG	11640
45	ACATCAACGT TTGGTGTGTG TAAATCATAG CCTATTTGAT GTTTCATGAT AAAAGTTCCA	11700
	TTTTCATCAT ATAAGACTGA CTTGGTACTC GTCGTTCCAA TGTCGACACC AATCATATAT	11760
	TTCATGATAA ATCCTTCTTT CTTTCATTTT AATTCAACCA AAATCCTTCA ATATCTTTAC	11820
50	CAACATCGTC GAAATTTAAA TGAAACGCTT CTTTCAAAAT TTGACTGTGC TATTGTTCCA	11880
	CTGCATCAAT AAACACTTGA TGATTATGAT GTATGCGTTC AAAATCTTGC GGGTTCTGTT	11940
55		

	AAAATGAGTT TAAATATTGA TGATTAGATG CTTTGATTAA TGTTCATGA AATTCAAAGT	12060
	CATGCTTCGT AAATGATTCT GCATCCTCAA ATTTTACTGC CACTTTCATC ATTTCAAGTT	12120
5	GTTTCTTCAT TTCTTTTACG ATAGGTAGTC GCTCTTGATT TTAACTCTT GAAAATGCAA	12180
	ATGACTCTAA CATCAGTCGC AAATCATACA TTTCTTTCTT TTCTTGTTCC CCAAACGGCA	12240
10	ACACATGTGC ACCCATTTCTT TCTAATTGGA TGAGTTGATT TTGTTGCAAT AATTTAAATG	12300
	CATCTCGAAT TGGCGAACGA CTCACATTAA ATTGCTTTGC CATTTGATTT TCAGTGAGTA	12360
	ACGTACCTTC AGCTATGTGA CCATTCACAA TGCCTAAGCG TAATTCTGCC GCGATACCTT	12420
15	CTCCAGTTGT CATACCTTCC AACCATTCTT CTGGATATCC ATACATCATC AAAGTCACTC	12480
	CTTCATTACA CGACATACTT GTATACAAGT ATGTTAATAT AGTTATTATG AGTTTGCAAG	12540
	CGCTTTCTTT ACGAGCACTA AAATAGTGAC CACCCCTTTT CGATTTAAT TTAAGGAAA	12600
20	TGGTCACTAT CACACGAATG ATTTAATTGT TATGTTGTAT GTGGGATATT TCTAATTGTT	12660
	CTGTACTCAT ATGCGCTTTA GGTACTTCAA TGCAATAATG CGTTTCATGA CAGTTTGAC	12720
	ATTCGAATCG ACGTGTGTG GCTGTATGTT TCGCTTTGAT AACTGCCAC AAAGATGGTG	12780
25	AGAATATATG CTGGCAGTTA GGACATAAAT AGGCAACCTT TTGTTGGTAA TAAAACTAA	12840
	CACCAATGCC ATAACCAATC ATAAATGGTA AAGCAATTAA AAACGGCCAT TTATTTTCA	12900
	TCAAAATTGC ACTTATAATG CTAGAATATT GAATTATTCC TATAATACCA GCACTAATCC	12960
30	AAATGTTACG ACGAATACTT TTCATTTTCA CTGATTTACT CATGACATGC TCTATGTCTT	13020
	TTAAGTGTGT GATTGGAGAC GTCGACGCTT CATTACGTA ATATTGAACA TTTTAAATT	13080
35	TGTTTAATAC CGCTTGTGTC TGTTTAACTT GTTGGTTAAT TTCTTGTTGT TTCATAGTTA	13140
	GTAAAGTATT GAGCGTCTTC AAAGTACCTT CACCTTTTAG CAACATATCT ATATCGCTTA	13200
	ACGCAACAACC TAAATCTTTA AGCAATAAGA TTAAGTCTAA TGTTTGTCGC TGTTGTTCTG	13260
40	TATACACACG ACGCTTTCCT TCTGTAAATC CTTGTGGTTT CAAAATACCT TTGCGATCAT	13320
	AATATTGAAT CGTTCGTGTT GTCACATTGC ATAAATTTGC GAGTTCTCCA GTCGAATAGT	13380
	TAGACATAGA TTCCACCTCC TATAATTACC ATAGTTGATG ACCCGACGTC ACGAGCAAGT	13440
45	ACAATTTCCA CATTTTAAAG AAATTTATTA TACTAGGCGT CTTATTTTTA TGATTTTCGTA	13500
	CCATGTTGAT TTACAACTC ACTCAAATA AGTAACACAC CTAATAACA TCTACTCTGT	13560
	TATTTAGAA TGAATTTGTT GTAATTTATC TTCAACTTCA GTAATCTCTG TCGCACATTC	13620
50	TTTCAGTAAA TCTCGATACT TTTCCGTCTC TGCAATGTTT TTATAACGTA TTTTATGTTT	13680
	TAAACTTGcC CACATATCCA TACCTATCGT TCTAATTGA ATTTCAACAG GCAATACCTC	13740
55		

## (2) INFORMATION FOR SEQ ID NO: 55:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1059 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

```

GGATAAGTTC AGGTAAATTC ATTCTTTTTT CAATTTTGAT TTTCATTGTT TCCGCCCTTT      60
TAAAATAAAG TTAGTTGCTT CTGTTCTCTCA TATTCCAAAT CACTTTGCTT TATATATGTT      120
TCAAGCTCTT CCGCTGTATC AAATGTCTTT TTCACACCTT GCCAACCTGG CACGATATGA      180
CCGTGAAAGT AATAAGTGCC ATTTACTACA TGGATATGTG CCACTCGTTC GTTATCCTGA      240
TACAGATATC TCTTAGATCC AAAGAATTGA TTTAGGTATT CTTTACCGGC GCTATCTGTC      300
ATGGTCATCA CTCCTTTTAA CAATTAGGCA GACCAAACGA CATGCATTCTG TCGTATAGCT      360
CTTCATTACT TATGCTTGCC TTATAGTTTT CAATCACATT GCTAACTTCT TTATGACTCA      420
TTGCTTTAAC TTGTTCTGCT GTATATTTTT CGCAGTCTTC TAATTCCAGT TGCTCCTGTA      480
ATGACATCAC ATATTCAACT TGTCTTTGGG TTGCCATCGT TAACCCCTCC ACAAGTCAAA      540
AGCTCTTTGG ACGTAAAACT TCGCCTTTGC TAAATCCTCA TGACCATTCT TTAACGGTGC      600
TCTAGACATG TATTTGATTG CATTACCTAT TGCGAATGCT AGTTGAGGTG GATACTGTGC      660
CGTAACCTGT TCGATAAAAT CTATAATTTT AATGTCGCCG TATGTGTAGT GCGCTGGTTG      720
CTTAACATTG TCTTGCGCTT CGTTCATATC TACTTTTCTG TTAGTGATTA CGCTCATTAT      780
GCTTCACTCC ATTTCTTGAA CATTTGGTTA TAAGTGACAT CGAACCAGTA CGGATCACGT      840
GAATGTTTTT GTGGCGTTCC ATCATAAAGC CATGGTCTTA ATCTTCTCTT TCTTTCCTGT      900
TCATATTCCG CTCTCACAAT TCGTTGGTAT CGGTTCAAAA TCGCTTTTTT TCTGATTTTT      960
TCTCTCCCTT TTTCTTCATC TTTnATtTGA CTCTnCATAT ATTCAACTTC TTCTGTAGAT     1020
nTTGAGTCCT TTCTTCACA CAATAATTCA nCGCCGCGC                                1059

```

## (2) INFORMATION FOR SEQ ID NO: 56:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30246 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

	GAAGTAAAAG AAGAATTAAA TTAAACATTA ACAATGGATG AAATTGAATA TGTCGGGACA	60
	ATTGTAGGTC CTGCATATCC ACAACAGGAT ATGTAACTG AGTTAAATGG ATTTGCGCA	120
5	TTAACCAAAA TCGATTGGGA AAACGTAACT ATCAATAATG AAATTACGGA TATACGCTGG	180
	ATTGATAAAG ATAATGATGC GTTGATTGCG CCTGCTGTCA AAGTTTGGAT TGAAACTTAT	240
10	GGTGGTAAAC ATGACAAATA ATGACACCAT CATGTTACGA CATTATGTCC CACAAGATTA	300
	TTGATGTTA GAAGCTTTTC AATTAAGTGA AAGTGATTG AAGTTTGTTA AAACGCCAGA	360
	GGAAAATATT ACAGCTGCAA TGTCTGATAA TGAAAGGTAT CCCATCGTTG TAATGGATGG	420
15	CAGGCAATGT GTGGCCTTTT TTACATTACA TCGTGAAAAA GGGGTCGCAC CATTTAGCGA	480
	TAACCAAGAT GCAGTATTTT TCAGGTCATT TAGTGTTGAT CAACGTTATC GTAATAGAGG	540
	AATAGGTAAA GTGGTAATGG AAAAATTGGC GTCATTATC ACTTCAACAT TTCAGGATAT	600
20	TAATGAGATT GTGTTAACGG TTAATACTGA CAATCCACAT GCCATGGCAC TTTATCGCCA	660
	ACAAGGATAT CAATATATGG GAGATAGTAT GTTCGTCGGA AGACCTGTTC ATATTATGGC	720
	GTAACTATA AAATAAATTA AATTTAAAAG CATCTTTACT CATCGTCGAC CACAACAATT	780
25	AATGATGAAT AAAGGTGCTT TTTGTTATAG ATCATCGGAC AATTTACTAT AGTAAAAAGC	840
	GACCTAGTGA ACAATTGACA TATATCCACA GGTCGCTTAA CTTAAGTTAT ATTGCTAGTT	900
30	GCGATTAATT GATAGACTCA TCATTTTTCG GCTGTCGAGA TGGTCTTTTT ATTA AAAATG	960
	CCGTAATCCA AGCCGTAATC GGAATACTGA TTGCAACGGC AATACCGCCT AAAATAATAG	1020
	AAATAAATTC TTGGGCAAAT ATTTTCGAGT TTATAATATG ACCAAATGAA TATTTAAGTT	1080
35	TGAAAAACCA AATAAATAAA GCAAGTTGGC CACCAAAAAA GGCAAGGTAA ATCGTGTTCG	1140
	CAGATGTCGC TAAAATTTCT CTACCAACAC GCATGCCAGA TTGGAATAAT TCGTATTGCG	1200
	TAACGTTgGA TTCACTTGAT GCAATTCATA AATGGGTGAA CTAATGGTAA TTGTTAAATC	1260
40	TATCAGAGCT GCAATAACAG CAAGAATAAT AGTGAACACC ATAAATTGAA CCATATCAAT	1320
	GCCAATATTC ATTGAATACA CATATGTTTC ATCTTGTTGT TCGGTTGaAA AGCCTGTAG	1380
	ATGACCGAAG TAGACCGATA AATAAATGAG TGTAATCAAC AATATTGTTG TAACGATAgT	1440
45	GctGgATAAA TGCaGCTTGT GTTTTAAAT TGTAATATT GAGTACGAAT AAATTACAAG	1500
	CGCCAATAAT AATGCAGAAA AAGAATGTGA CGACATAAAT CGGTACGCCA AAAATAATCA	1560
	ATACAATACT AATAATTAAA ATAGCGAAAT TAAAAAATAG GGTAAATAA GAGATGAATC	1620
50	CCTTTTACC TCCGAAAATT ATCATCAGAA AGAGGAGCAA TAACGCCAAT ATAAATACAG	1680
	CATTCAATTGT TTCGCCCTCC TTAATGTTTC AAATATTTCC ATAAACAATA TTGTGATAGG	1740
55		

## EP 0 786 519 A2

	CATCGAAATA GTATAAGTCA CTGTATTGGC ATTTTTTAAA AAGATTAAAA ACATAGGTAG	1860
	TGCACCGGAT AAATATGAGA ATAATAAGAT GTTAGTCATT GTTCCCATAA TATCTTGGCC	1920
5	GATGTTTCGC CCAGCAAGCG CCCATCTCCT CATTGAAATG TGTGGCGTAC GCTGTAAAAT	1980
	TTTCATGCATA CCACTAGCAA TTGTAATTGC AACATCCATA ATAGCGCCAA GTGAACCTAT	2040
	TAACACTGAG GCTAGGAAGA TATCTTTCGG TGGTAATGAT AAAAAGTTCA TCGTTTCATA	2100
10	TTTAATGCCT TTACCATCTG TCATATATAT GATTAATTCT GTTAAACCTA TACTCAAAAA	2160
	AGTTCCGATA ATTGTACTGG CTATGGTAAT GAGTGTACGC ATATGCCAGC CTGTAACGAG	2220
15	CAATAAAGTG AGTATTGTTG AACAGATCAT GGCAATGGTC ATGAGTAAGA ATAAATTAAT	2280
	ATTGCTATGT TGAATATGAA TGTAATTCG GATTAATATG GCAATAGAAT TCAAGATTAA	2340
	CGATAAAATC GATTGCAGTC CGACTTTGCG ACCAACCAAT AATACAGTTA ATAAGAACAA	2400
20	ACCAGTGATG ATAACCGTTA AGGTATCACG CTTCTTTTCT ATAATATAAG CATCACTCGG	2460
	CTTGTTAGAA ATATGTAATA ATACTTTTTC GTGTGTGCGA AATGCCTCAG AATCTGCTTG	2520
	CGATTTGACG TACTGATGAT TAATCGTCGT CGTTTCTCCA GCAAATTGAC CATTTAATAT	2580
25	TTTGACTTTT AATTGATTTT TATATTTAAT ATCACGATTA TTTTGTGCAT CTTTGTAGG	2640
	TGTCGAAGAA ACATGTTTGA CATCTATAAT TTGACCAATT GGTGTGTTGT AAAAGTTCTC	2700
	ATTATTGAAT GTAAATAAAA TAGCACCAAT GAATGCGATG CAGAACAAAC CTAAAAATTAT	2760
30	ATTAAATGGC TTTGTAAATA AATTTCTATA TTTCAAAAAC AAAACCCCAA TTCTATGAAT	2820
	GAATTAATAT GGTGATTATA CGCCCTTAAT TTTTATTTT CAAAGATATT ACTGCTAAGT	2880
	GTA AACGAA AATCATCATT GATAGCATCG AATTACTTAA TGGAATGTAG ACGTTTTAGT	2940
35	CATTAAATTGC TGAATAAGTG TTAATAATAT GCCAATATCA CTCTTTGTAT AAGGCTCCTT	3000
	TGTAATAGCA CATATCGTTC TTTTAAATC AGTATGATCT AATTTTATAT CTATCCATGA	3060
40	TTTAGATTCT GGTAAATGTA TATTTTGTGA TGAAATGATG TAACCTTCTT TTTGACGAAG	3120
	GAGATACGTC GCAAGTGGTT GGCTACTGAT TGTGTATACA TCTGATTAG TAATCTTGCG	3180
	CAATTGTTTT TTTACAGTTT CGGCAAATGG TGCCAAGCAA TAAATATGAC TATGCTCAAA	3240
45	CTGAATTAAT GGTGGGTGTG TCGCCATCGT AATTGGATCG TCTGAAGGCG CATATAAATG	3300
	ATAGTGCTCT TCGAATAAAG GTAGCATATG TAATTGTTG TGTTTACGTA TTTCTGGTGT	3360
	AAGTTCCGTG AAACCAATGT CTATATCCC ATTTAATACG CTATTTATAA TTGTGTCATG	3420
50	TTCTAATAAG CTCGGTATGA CATGTGTATC ATTTTGTAAA TGAAACGTTT GGATAAGTGG	3480
	TAGTAACATG TGGGATACGT CACTCTCATC ATAGCCAATG TAGATACTTT TATTTTGTAGT	3540

55

TTCATTAAAT AATAATTTCC CTTGAGATGT GAGCGTAATA TTGCGTCCTT GCTTTTAAAA 3660  
 TAAAGACACA ITAAGTTCTT GTTCTAATAA TGTAATTTGA CGGCTTATCG CTGATTGAGC 3720  
 5 AATGTTTAGT ICAAGTGCTG TTTGCGAGAT ATGTTCTCTT TTAGCGACCT CGATAAAATA 3780  
 TCTTAATTGT ITAATTTCCA TAGCGATATA GGCACCTCCA AAAATGAGTG TTTTGTAAC T 3840  
 ATTATAGCAA TATTATTGAT AAAIGTTCTA TTTTITAGAT GAATATCTTC TATTTTATAT 3900  
 10 ATTGAACAGA TAAATTTTTT AGATTATAGT AATTATCATT AATAACTAAT ATCAGAATAT 3960  
 TCTAAAAAAG GGGTGTGCAT CATGCACAAT GAGAAATTAA TTAAAGGCTT ATATGACTAT 4020  
 15 CGTGAGGAAC ATGATGCGTG TGGTATTGGT TTTTATGCGA ATATGGATAA TAAAAGGTCT 4080  
 CACGACATCA TTGATAAATC GCTTGAAATG TTGCGACGCT TAGATCACAG GGGCGGGGTC 4140  
 GCGCGAGATG GCATCACTGG TGATGGCGCA GGTATTATGA CTGAAATACC TTTTGCATTT 4200  
 20 TTCAAACAAC ATGTAACGGA CTTTGATATC CCAGGTGAAG GTGAATATGC CGTGGGGTTA 4260  
 TTTTITTTCCA AAGAACGCAT TTTAGGTTCT GAACATGAAG TAGTTTTTAA AAAATATTTT 4320  
 GAAGGCGAAG GGTTATCAAT TCTTGGTTAT CGTAATGTAC CAGTTAATAA AGATGCCATT 4380  
 25 GCTAAACATG TAGCAGATAC GATGCCAGTC ATTCAACAAG TGTATTATGA TATTAGGGAC 4440  
 ATTGAAGATG TTGAAAAGCG TTTGTTTTTA GCGAGAAAAC AATTAGAGTT CTATTCGACT 4500  
 CAGTGCGATT TAGAATTGTA TTTTACGAGC TTATCACGCA AAACAATTGT ATATAAAGGT 4560  
 30 TGGTTACGAT CAGACCAAAT TAAAAAATA TATACAGATT TATCGGATGA TTTATATCAA 4620  
 TCAAAGCTAG GGTTAGTGCA TTCGAGATTT AGTACGAATA CATTCCCGAG TTGGAAAAGG 4680  
 GCACATCCTA ACCGTATGTT AATGCATAAT GGTGAGATTA ACACGATTAA AGGTAATGTA 4740  
 35 AACTGGATGC GAGCACGCCA ACATAAATTA ATCGAAACAT TATTTGGCGA GGATCAACAT 4800  
 AAAGTGTTC AAATTGTGCA TGAGGATGGT AGTGACTCTG CCATTGTAGA TAATGCGCTA 4860  
 GAGTTCTTAT CGTTAGCCAT GGAGCCAGAA AAGGCAGCGA TGTTACTCAT ACCTGAACCT 4920  
 40 TGGTTATATA ATGAAGCGAA TGATGCAAAT GTACGTGCGT TTTATGAATT TTATAGTTAT 4980  
 TTAATGGAAC CGTGGGATGG TCCTACAATG ATTCGTTCT GTAACGGTGA CAAACTTGGC 5040  
 GCGCTTACAG ATAGAAATGG ATTACGTCCA GGTGTTTATA CGATTACTAA AGATAACTTT 5100  
 45 ATTGTCTTTT CATCTGAAGT GGGTGTGTG GACGTACCTG AAAGTAATGT TGCTTTTAAA 5160  
 GGTCAATTGA ATCCTGGAAA GTTATTGCTT GTTGATTTTA AACAGAATAA AGTCATTGAA 5220  
 50 AATAATGATT TAAAAGGTGC GATTGCTGGA GAATTACCAT ATAAAGCGTG GATTGATAAC 5280  
 CATAAAGTTG ACTTTGATTT TGAAAAATA CAATATCAAG ATTCGCAATG GAAAGATGAG 5340

55

	CAGGAACTTG TAGAAGGTAA GAAGGATCCT ATCGGTGCAA TGGGATATGA TGCGCCAATT	5460
	GCAGTGTGTA ACGAGCGACC AGAATCACTA TTTAATTACT TTAAACAGCT GTTTGCACAA	5520
5	GTTACGAATC CACCAATTGA TCGGTATCGT GAAAAAATCG TAACGAGTGA ACTTTCTTAT	5580
	TTAGGTGGCG AAGGTAACCT ACTAGCACCT GACGAAACGG TTTTAGATCG TATTCAATTG	5640
10	AAAAGGCCGG TATTGAATGA ATCACACTTA GCAGCGATTG ATCAGGAACA TTTTAAATTA	5700
	ACTTATTAT CAACGGTATA TGAAGGGGAT TTGGAAGATG CGTTAGAAGC ATTAGGCCGA	5760
	GAAGCAGTGA ATGCTGTAAA GCAAGGCGCT CAAATTCTAG TGTTAGATGA TAGTGGATTA	5820
15	GTTGATAGCA ATGGCTTTGC AATGCCGATG TTAATCGCAA TAAGTCATGT GCATCAATTA	5880
	CTTATTAAAG CAGATTTACG TATGTCTACA AGTTTAGTCG CTAAATCTGG TGAGACACGA	5940
	GAAGTGCATC ATGTTGCTTG TTTACTCGCA TATGGCGCGA ATGCAATTGT GCCATACCTA	6000
20	GCGCAACGTA CAGTTGAACA ACTGACATTG ACAGAAGGGT TACAAGGCAC CGTTGTCGAT	6060
	AATGTTAAGA CATATACGGA TGTATTGTCA GAAGGTGTCA TTAAAGTAAT GGCTAAGATG	6120
	GGAATTTGCA CAGTGCAAAG TTATCAAGGG GCACAAATAT TTGAAGCGAT TGGCTTGTCT	6180
25	CATGATGTGA TTGATCGTTA TTTTACTGGG ACACAGTCTA AGTTATCTGG TATTTGATT	6240
	GATCAAATTG ATGCTGAAAA TAAAGCACGT CAACAAAGTG ATGATAATTA TCTTGCATCA	6300
	GGTAGTACAT TCCAATGGAG ACAACAAGGT CAACATCATG CTTTTAATCC GGAATCTATT	6360
30	TTCTTATTGC AGCACGCATG TAAAGAAAAT GACTATGCGC AATTTAAAGC ATACTCTGAA	6420
	GCGGTGAACA AAAATAGAAC AGATCACATT AGACATTTAC TTGAATTTAA AGCATGTACA	6480
	CCGATTGACA TCGACCAAGT TGAACCGGTA AGTGACATTG TCAAACGCTT TAATACAGGG	6540
35	GCGATGAGTT ATGGATCGAT TTCAGCGGAA GCACATGAAA CGTTAGCACA AGCCATGAAC	6600
	CAATTAGGTG GAAAGAGTAA TAGTGGTGAA GGTGGCGAAG ATGCAAAACG TTATGAAGTA	6660
40	CAAGTTGATG GAAGCAACAA AGTAAGTGCG ATTAAACAAG TTGCTTCTGG GCGTTTTGGT	6720
	GTAAC TAGTG ATTATTTACA ACATGCCAAA GAAATTCAAA TTAAAGTTGC GCAAGGTGCA	6780
	AAGCCTGGTG AAGGTGGTCA ATTACCTGGT ACTAAGGTAT ATCCGTGGAT TGCGAAGACA	6840
45	AGAGGGTCAA CGCCAGGTAT CGGTCTGATT TCACCACCGC CACATCATGA TATTTATTCA	6900
	ATAGAAGATT TAGCGCAACT GATACATGAT TTGAAAAATG CGAATAAAGA TGCAGATATC	6960
	GCGGTAAAT TAGTTTCGAA AACAGGTGTT GGTACCATTG CATCTGGGGT GGCAAAAGCA	7020
50	TTTGAGATA AAATTGTCTAT CAGTGGTTAC GATGGTGGTA CAGGGGCTTC ACCCAAAACG	7080
	AGTATTCAGC ATGCCGGTGT TCCTTGGGAG ATTGGTTTAG CAGAAACACA TCAAACATTA	7140

55

AAAGATGTAG CGTACGCATG TCGCTTGGG GCGGAAGAAT TTGGATTTGC AACTGCACCA 7260  
 TTAGTGGTGT TGGGCTGTAT TATGATGCGT GTATGCCATA AAGATACATG TCCAGTAGGA 7320  
 5 GTTGCAACTC AAAACAAAGA TTTACGTGCT TTATATAGAG GTAAAGCACA TCATGTTGTT 7380  
 AATTTTATGC ATTTTATGTC ACAAGAATTA AGAGAAATTT TAGCATCTTT AGGTTTGAAA 7440  
 CGTGTAGAAG ACTTAGTTGG AAGAAGTCAT TTATTACAAC GATCATCAAC ATTAAGGCG 7500  
 10 AATAGCAAAG CGGCTAGTAT TGATGTTGAA AACTGTTAT GTCCTTTCGA TGGGCCAAAC 7560  
 ACAAAGAAA TTCAACAAAA TCATAATCTT GAGCATGGAT TTGATTAAAC AAATTTATAT 7620  
 GAAGTAACGA AGCCATATAT TGCTGAAGGG CGTCGCTATA CAGGTAGCTT TACAGTAAAT 7680  
 15 AATGAACAAC GTGATGTAGG GGTATTACA GGTAGTGAGA TTTCGAAACA ATATGGAGAA 7740  
 GCAGGACTTC CTGAAAATAC AATTAATGTT TATACGAATG GTCATGCTGG TCAAAGTCTT 7800  
 GCAGCATATG CACCGAAAGG CTTAATGATT CATCATACTG GAGATGCGAA TGAATATGTT 7860  
 20 GGTAAAGGAT TATCTGGTGG TACGGTCATT GTCAAAGCAC CTTTGAAGA ACGACAAAAT 7920  
 GAAATTATTG CTGGTAACGT CTCATTCTAT GGTGCGACAA GTGGTAAGGC ATTTATTAAC 7980  
 25 GGTAGTGCAG GAGAAAGATT CTGTATTAGA AATAGTGGTG TAGATGTTGT CGTTGAAGGT 8040  
 ATCGGCGACC ATGGATTAGA GTATATGACT GGTGGACATG TCATTAATTT AGGTGATGTA 8100  
 GGTAAGAACT TCGGTCAAGG TATGAGTGGT GGTATTGCTT ACGTTATCCC GTCTGATGTA 8160  
 30 GAAGCTTTTG TTGAAAATAA TCAACTAGAT ACGCTTTCGT TTACAAAGAT TAAACACCAA 8220  
 GAAGAAAAAG CATTCTATTA GCAATGCTG GAAGAACATG TGTCACACAC GAATAGTACG 8280  
 AGAGCGATTG ATGTGTTAAA ACATTTTGAT CGCATTGAAG ATGTCGTCGT TAAAGTTATT 8340  
 35 CCTAAAGATT ATCAATTAAT GATGCAAAAA ATTCATTTGC ACAAATCATT ACATGACAAT 8400  
 GAAGATGAAG CGATGTTAGC TGCAATTTAC GATGACAGTA AAACAATCGA TGCTAAACAT 8460  
 AAACCAGCCG TTGTGTATTA AGGAAAGGGG GAGATACGAT GGGTGAATTT AAAGGATTTA 8520  
 40 TGAAGTATGA CAAACAGTAC TTAGGTGAAT TATCACTGGT AGACCGTTTG AAGCATCATA 8580  
 AAGCATATCA ACAACGATTT ACTAAAGAAG ATGCCTCTAT CCAAGGTGCA CGATGTATGG 8640  
 ATTGTGGAAC GCCGTTTGT CAAACCGGAC AACAGTATGG TAGGGAAACA ATAGGTTGTC 8700  
 45 CAATTGGAAG CTACATTCCT GAATGGAACG ACTTAGTGTA TCATCAAGAT TTTAAACTG 8760  
 CTTATGAACG CTTAAGCGAA ACAAATAACT TTCCTGACTT TACAGGGCGT GTATGTCCTG 8820  
 CACCATGCGA AAGTGCTTGT GTGATGAAGA TTAATAGAGA ATCGATTGCG ATTAAGGTA 8880  
 50 TTGAACGCAC AATTATTGAT GAAGCTTTTG AAAATGGTTG GGTAGCGCCG AAAGTTCCGA 8940

55



	CTGAAGAACT TAATCTACTA GGATATCAAG TAACTATTTA TGAACGTGCT AGAGAATCAG	9060
	GCGGTTTATT AATGTATGGT ATTCCGAATA TGAAACTTGA TAAAGATGTG GTTCGACGTC	9120
5	GTATTAAGTT AATGGAAGAA GCGGGCATT CTTTCATTAA TGGTGTGAA GTCGGTGTG	9180
	ATATTGATAA AGCAACGTTA GAATCTGAGT ATGATGCCAT TATATTATGT ACTGGTGAC	9240
	AAAAAGGTAG AGATTTACCT TTAGAAGGAC GCATGGGTGA TGGTATACAT TTCGCTATGG	9300
10	ATTATTTAAC TGAACAAACG CAGTTGTAA ATGGAGAAAT TGATGATATA ACAATAACTG	9360
	CAAAAGATAA GAATGTCATT ATCATTGGTG CTGGTGATAC AGGGGCAGAC TGTGTAGCGA	9420
	CAGCATTAAAG AGAAAATTGT AAATCGATTG TTCAATTTAA TAAATATACG AAATTGCCAG	9480
15	AAGCAATTAC ATTTACAGAA AATGCATCAT GGCCTTTAGC AATGCCGGTG TTTAAAATGG	9540
	ACTATGCGCA CCAAGAGTAC GAAGCTAAGT TTGGTAAGGA ACCACGTGCA TATGGTGTTT	9600
20	AAACAATGCG TTACGATGTT GACGATAAAG GACACATACG TGGTTTGTAT ACTCAAATTT	9660
	TAGAGCAAGG CGAAAATGGT ATGGTCATGA AAGAAGGACC TGAAAGATTT TGGCTGCTG	9720
	ACCTTGATTT ATTATCAATC GGCTTCGAAG GTACAGAACC AACAGTACCG AATGCTTTTA	9780
25	ACATTAAAC GGATAGAAAT CGAATCGTG CGGATGATAC AACTATCAA ACTAATAATG	9840
	AAAAGGTATT TGCTGCTGGA GATGCTAGAC GTGGTCAAAG TTTAGTTGTA TGGGCAATTA	9900
	AAGAAGGTAG AGGCGTAGCG AAAGCAGTAG ATCAGTATTT AGCTAGTAAA GTTTGTGTAT	9960
30	AATCTTGTA TGGAAATGGT GGTACGTTG ACGTTGTGAC ATGCTGAATC GAGTTTGAAA	10020
	AAATCTAGTA TCTATCAACG TCACATGCCA TCTTTGTAAC CTAAAAACAA AGTTTGTA	10080
	GACAACAAAT AGATTAATTA TAAGTAGTGA TTTTACAT TCGTTTATAG GTCAACTGTA	10140
35	GTGGAAGACA ATGATTTGTG GTAATCATGT AATGCTTAAA AACAATATTG ACTTTTACAG	10200
	AACGTTTATA TATGATAAAT ATTGTTGTTA GGAGGAATAC CCAAGTCCGG CTGAAGGGAT	10260
	CGGTCTTGAA AACCAGACAGG GGCTTAACGG CTCGCGGGG TTCGAATCCC TCTTCCTCCG	10320
40	CCATCAATAT TTATATTAAA TTCTATATAT AATGAAGGTA AGTGCTCAAA TTTTGAGTAT	10380
	TTACCTTTTT TATTTGTCTT TGAATGGCTC GTAATTTTTG ATAATAGAAA TGATAAGGCA	10440
	TTGAGATTGG AAGGGCATTT GGCTTGTGCA ATATACATAG CTAAATGTCT TTTTGTGTTT	10500
45	GTGAAATATG ATGGATGGCT TGTGTGGACA AGTTTGCTAT TTATAGATAT GCATTTTCA	10560
	ATTTAGGAGT TGGCCATGCA TCTACACTTT ATAATGGTGA GAGCGTGGTG AGGTATTGTT	10620
50	AATAACGCAA TTGTAGCGAG GAGTTATGTC TACATATGTC GTTATGGCTC ATTGATTTTC	10680
	TGAAATGGCT ACCCCAGATA ATTGTGACAA AATAAAAATA TTTTGTGAA AGCCTTTACA	10740

55

	TAAAAAGAGA AGATGTAAAA GCCATCGTAA CCGCTATTGG GGGAAAAGAA AATCTTGAAG	10860
	CTGCAACGCA TTGTGTAACA CGATTACGTT TAGTGCTGAA GGATGAAAGT AAAGTTGATA	10920
5	AAGACGCATT AAGTAATAAC GCGTTGGTCA AGGGGCAGTT TAAAGCAGAC CATCAATATC	10980
	AAATTGTCAT TGGTCCAGGA ACAGTCGATG AAGTGTATAA GCAGTTTATT GATGAAACAG	11040
	GTGCTCAAGA AGCTTCGAAA GATGAAGCGA AACAAGCAGC TGCACAAAAA GGAATCCAG	11100
10	TACAACGTTT GATCAAATTG TGGGGGATA TTTTATACC AATATTACCT GCGATTGTGA	11160
	CAGCTGGTTT GTTAATGGGA ATCAATAATT TACTTACAAT GAAAGGTTTA TTTGGTCCAA	11220
	AAGCACTTAT TGAGATGTAT CCACAAATTG CTGATATTTC AAACATCATT AATGTGATTG	11280
15	CGAGTACGGC ATTTATTTTC TTACCAGCAT TAATTGGTTG GAGTAGTATG CGTGTATTTG	11340
	GTGGTAGTCC GATTCTAGGC ATAGTCTTAG GTTTGATTTT AATGCATCCG CAATTAGTAT	11400
	CTCAGTATGA TTTGGCAAAA GGAATATTTC CGACGTGGAA CTTATTTGGC TTAGAGATTA	11460
20	AGCAGTTGAA TTACCAAGGT CAAGTGTTGC CAGTTTAAAT TGCAGCTTAC GTTCTAGCTA	11520
	AAATTGAAAA AGGATTAAAT AAAGTCGTTT ACGATTGAT AAAAATGTTG GTCGTTGGAC	11580
25	CCGTAGCGCT TTTAGTTACT GGATTTTTAG CATTTATTAT CATGGACCA GTTGCCTTAT	11640
	TGATTGGTAC AGGTATTACA TCTGGTGTTA CATTTATATT CCAACATGCA GGATGGCTTG	11700
	GCGGAGCAAT ATATGGATTG TTATATGCAC CACTTGTAAT TACAGGACTA CACCATATGT	11760
30	TTTTAGCAGT AGATTTCCAA TTGATGGGTA GCAGCTTAGG CGGTACGTAT TTATGGCCAA	11820
	TTGTTGCGAT TTCCAATATT TGTCAGGGCT CTGCAGCATT TGGAGCATGG TTTGTCTATA	11880
	AACGTCGTAA AATGGTTAAA GAAGAAGGCT TGGCATTAAAC ATCTTGTATT TCTGGTATGT	11940
35	TAGGTGTTAC TGAACCAGCC ATGTTCCGGT TGAACCTACC TCTGAAATAT CCATTTATCG	12000
	CTGGGATATC AACGTCCTGT GTATTGGGGG CAATCGTTGG TATGAATAAC GTACTTGGAA	12060
	AAGTTGGTGT TGGTGGCGTG CCAGCATTCA TTTCAATTCA AAAAGAATTT TGGCCAGTAT	12120
40	ATCTTATTGT GACAGCTATT GCTATTGTTG TACCATGTAT ACTAACAATT GTGATGTCTC	12180
	ATTTTAGTAA ACAAAAAGCG AAAGAAATTG TTGAAGATTA ATAAAATAAA AAAGGGCGT	12240
	TCGTTATTTG GACGTCCTTT ATTACGTTAT AAGGTGGTAA TTGTGTGTCG AAAGAAATAG	12300
45	ATTGGAGAAA ATCCGTTGTA TATCAAATTT ATCCTAAGTC GTTTAATGAT ACGACGGGGA	12360
	ATGGTATAGG AGATATCAAT GGAATTATAG AAAAATTGGA TTATATCAAG TTATTGGGTG	12420
50	TTGATTATAT TTGGTTAACA CCAGTGTATG AATCACCAGT GAATGATAAT GGCTATGATA	12480
	TCAGCAATTA TTTAGAAATC aATGAAGACT TTGGAACGAT GGATGATTTT GaAAAGTTAA	12540

55

	CGACCGAGCA TGaATGGTTT AAAGAAGCCC GTAAATCTAA AGATAACCCy TATAGAGATT	12660
	ATTACTTTTT CAGATCATCT GAAGACGGGC CGCCAACAAA TTGGCATTCT AAATTCGGTG	12720
5	GTAATGCATG GAAGTATGAT TCTGAGACAG ATGAATATTA TTTACATTTA TTTGATGTCA	12780
	GTCAAGCTGA TTAAATTGG GATAATCCGG AAGTACGTCA ATCGTTATAT CGCATAGTCA	12840
	ATCATTGGAT AGACTTCGGC GTTGATGGTT TCGATTGA TGTCATTAAC TTAATTTCTA	12900
10	AAGGTGAATT TAAGGACTCT GACAAAATAG GTAAAGAATT TTATACGGAT GGTCCTAGAG	12960
	TGCATGAGTT TCTGCATGAA TTAAATCGTC AAACGTTTGG TAACACTGAC ATGATGACTA	13020
	TAGGAGAAAT GTCTTCGACG ACGATTGAAA ATTGTATTAA GTATACACAA CCAGAACGCC	13080
15	AAGAATTGAA TAGTGTTTTT AATTTTCATC ATCTAAAGGT TGATTATGTT GATGGTGAAA	13140
	AGTGGACAAA TGCGAgcTTG nATTTTCATA AGTTAAAGGA AATTCTGATG CAATGGCAAC	13200
	GAGGTATTTA TGACGGTGGC GGATGGAACG CGATTTTCTG GTGTAATCAT GATCAGCCAC	13260
20	GGGTAGTGTC TAGATTGGT GATGATACGT CGGAAGAGAT GAGGATACAA AGTGCTAAAA	13320
	TGTTAGCTAT CGCACTGCAT ATGTTGCAAG GGACGCCATA TATTTACCAA GGTGAAGAAA	13380
25	TTGGTATGAC GGACCCACAT TTTACATCAA TAGCACAATA TCGTGATGTT GAATCGATTA	13440
	ATGCCTACCA TCAGTTGTTA AGTGAAGGGC ATGCTGAAGC GGATGTGTTA GCGATTTTAG	13500
	GACAGAAGTC ACGAGACAAT TCGAGAACGC CTATGCAATG GAGTGATGAT GTTAATGCTG	13560
30	GATTTACAGC TGGTAAnCCT TGGATTGATA TTTCGGAAAA TTATCATCAG GTCAACGTTA	13620
	GACAAGCACT TCAGAATAAA GAGTCTATTT TCTATACGTA TCAAAAATTA ATACAATTAA	13680
	GACATACGCA TGATATTATT ACGTATGGAG ACATTGTGCC ACGTTTATG GATCATGATC	13740
35	ATTTATTTGT TTATGAACGT CATTATAAGA ATCAACAATG GCTAGTAATT GCGAATTTCT	13800
	CAGCAtCGGC TGTTGATTTG CCAGAAGGAT TGGCTAGAGA AGGTTGTGTT GTGATTCAA	13860
	CAGGCACAGT GGAAAATAAT ACGATAAGCG GGTGTTGGTGC AATTGTAATC GAAACAAACG	13920
40	CGTAAATAA ATTGAGTGGG TCGGTTTATA TCGCGAAACA AAAAAAGTTT ATGAAGATTT	13980
	ATGAGGCGTT GAAAGAAGAT ATATTAAACG GGCAGATTCA ATATGGTGAA CAAATTCCGT	14040
	CTGAACATGA TTTGGTGCAA TTGTACCAGT CATCTCGAGA GACCGTCCGT AAGGCATTAG	14100
45	ATTTGTTGGC ATTAGACGGC ATGATTCAAA AGATTCATGG TAAAGGGTCA CTTGTCATTT	14160
	ATCAGGAGGT TACAGAGTTT CCATTTTCTG AACTTGTTAG TTTTAAAGAA ATGCAAGAAG	14220
50	AAATGGGCGT CGCATATTTA ACTGAAGTTG TTGTGAATGA GGTGTTGAA GCGCATGAAG	14280
	TTCCAGAAGT TCAACATGCT TTAACATCA ATTCTAGTGA ATCACTCATT CATATTGTTA	14340

55

	TTGTTTCAGA TATAGGTAAT GATGTTGCCA GTGATTCTAT TTATGATTAT TTGGAAAAGG	14460
	TATTAAATCT TAATATTAGT TATTCAAGTA AGTCTATTAC TTTTGAACCG TTTGATGAAC	14520
5	AAGCATATCA ATTGTTTGGT GATGTATCGG TGGCTTATTC AGCAACAGTT CGAAGTATTG	14580
	TGTATTTAGA AAATACAATG CCGTTTCAAT ATAATATTTT AAAACATCTT GCAAATGAAT	14640
	TTAAATTTAA TGACTTCTCA AGACGTCGTA TAAAGTAAAC AATGATATAA ATGATTTATA	14700
10	CTTGCAATTA ACTATTAAAA TATAGTAATA TATATCTTGC CGTGCTAGGT GGGGAGGTAG	14760
	CGGTTCCCTG TACTCGAAAT CCGCTTTATG CGAGGCTTAA TTCCTTTGTT GAGGCCGTAT	14820
	TTTTCGGAAG TCTGCCCAAA GCACGTAGTG TTTGAAGATT TCGGTCCTAT GCAATATGAA	14880
15	CCCATGAACC ATGTCAGGTC CTGACGGAAG CAGCATTAAAG TGGATCATCA TATGTCCCGT	14940
	AGGgTAGCCG AGATTTAGCT AACGACTTTG GTTACGTTCTG TGAATTACGT TCGATGCTTA	15000
	GGTGACGGT TTTTTATTTT TTAAATATTA AACCGATTAT TAAGAGTTGA AAATATATAA	15060
20	TTATAGAAGC TACTTTCTTG AAGACAATTC AGCGTATTAT ACGTGGAACA TGTTTGTGGG	15120
	AAGTAGCTTT TTTATATGTG AAGTTTGATT CAAGTGAAC TCGATGTGCAG TTTGAATGAT	15180
	TTTTGTGTCA ATGAAAAGTA AGAAGTTATA ATTTGATGAT AAAGAAATGA TGGTGAAATG	15240
25	AGGGGGAGTA TCTTACAATA GAATTATTAA TGAGATACGT TATGATTATT GACAATCAAA	15300
	TGCCTACGGA GGACATATGC AAATATATTT AAGTACTTTA ACAGAGTTAG ATTATGATAA	15360
	ATCTTTAAAT AGTATTGAAG AAAGTTTGA TGATAATCCT GAAACGAGTT GGCAAGCACG	15420
30	TGCGAAAGTA AAACATTTAA GAAAATCTCC TTGCTATAAT TTTGAATTAG AAGTAATAGC	15480
	GAAAAATGAA AATAACGATG TCGTTGGACA CGTTTTATTA ATTGAAGTAG AAATTAATAG	15540
35	TGATGATAAG ACGTATTATG GTTTGGCGAT TGCCTCTTTA TCAGTTCATC CTGAATTACG	15600
	TGGA <del>C</del> AAAAA TTAGGTCGTG GCTTGGTTCA AGCAGTAGAA GAGCGTGCCA AAGCACAAGA	15660
	GTATAGTACG GTTGTTGTAG ACCATTGTTT TGA <del>C</del> TACTTT GAAAAGTTGG GTTATCAAAA	15720
40	TGCTGCTGAG CATGACATTA AATTAGAATC TGGTGATGCA CCGTTACTTG TAAAATATTT	15780
	ATGGGATAAT TTGACGGATG CACCACACGG AATCGTAAAA TTTCCAGAAC ATTTTTATTA	15840
	ATTGTTCAAT TAAGAAGTAA AGGTATTATC ATGCTATAAT GAGAGGTAAT TGTTTATGGA	15900
45	GGTGCTAACT TGAATTATCA AGCCTTATAT CGTATGTACA GACCCCAAAG TTTCGAGGAT	15960
	GTCGTCGGAC AAGAACATGT CACGAAGACA TTGCGCAATG CGATTTTCGA AGAAAAACAG	16020
	TCGCATGCTT ATATTTTTAG TGGTCCGAGA GGTACGGGGA AAACGAGTAT TGCCAAAGTG	16080
50	TTTG <del>C</del> TAAAG CAATCAACTG TCTAAATAGC ACTGATGGAG AACCTTGTA TGAATGTCAT	16140

55

AATAATGGCG TTGATGAAAT AAGAAATATT AGAGACAAAG TTAAATATGC ACCAAGTGAA 16260  
 TCGAAATATA AAGTTTATAT TATAGATGAG GTGCACATGC TAACAACAGG TGCTTTTAAT 16320  
 5 GCCCTTTTAA AGACGTTAGA AGAACCTCCA GCACACGCTA TTTTATATT GGCAACGACA 16380  
 GAACCACATA AAATCCCTCC AACAATCATT TCTAGGGCAC AACGTTTGA TTTTAAAGCA 16440  
 ATTAGCCTAG ATCAAATTGT TGAACGTTTA AAATTTGTAG CAGATGCACA ACAAATTGAA 16500  
 10 TGTGAAGATG AAGCCTTGGC ATTTAtcgCT AAAGCGTCTG AAGGGGGTAT GCGTGATGCA 16560  
 ITAAGTATTA TGGATCAGGC TATTGCATTT GGTGATGGTA CGTTAACATT GCAAGATGCG 16620  
 TTGAATGTCA CAGGTAGCGT ACATGATGAA GCGTTGGATC ACTTGTTTGA TGATATTGTA 16680  
 15 CAAGGTGACG TACAAGCATC TTTTAAAAAA TACCATCAGT TTATAACAGA AGGTAAAGAA 16740  
 GTGAATCGCC TAATAATGa TATGATTTAT TTGTGCaGAG ATACGATTAT GAATAAAACA 16800  
 TCTGAGAAAG ATACTGAGTA TCGAGCACTG ATGAACCTAG AATTAGATAT GTTATATCAA 16860  
 20 ATGATTGATC TTATTAATGA TACATTAGTG TCGATTCTGT TTAGTGTGAA TCAAAACGTT 16920  
 CATTTTGAAG TGTTGTTAGT AAAATTAGCT GAGCAGATTA AGGGTCAACC ACAAGTGATT 16980  
 25 GCGAATGTAG CTGAACCAGC ACAAATTGCT TCATCGCCAA ACACAGATGT ATTGTTGCAA 17040  
 CGTATGGAAC AGTTAGAGCA AGAACTAAAA ACACTAAAAG CACAAGGAGT GAGTGTGCGT 17100  
 CCTGTTCAAA AATCTTCGAA AAAGCCTGCG AGAGGCATAC AAAAATCTAA AAATGCATTT 17160  
 30 TCAATGCAAC AAATTGCAAA AGTGCTAGAT AAAGCGAATA AGGCAGATAT CAAATTGTTG 17220  
 AAAGATCATT GGCAAGAAGT GATTGATCAT GCCAAAAATA ATGATAAAAA ATCACTCGTT 17280  
 AGTTTATTGC AAAATTCCGA ACCTGTGGCG GCAAGTGAAG ATCACGTACT TGTGAAATTT 17340  
 35 GAGGAAGAGA TCCATTGTGA AATCGTCAAT AAAGACGACG AGAAACGTAG TAGTATAGAA 17400  
 AGTGTGTAT GTAATATCGT TAATAAAAAAC GTTAAAGTTG TTGGTGTACC ATCAGATCAA 17460  
 TGGCAAAGAG TTCGAACGGA ATATTTACAA AATCGTAAAA ACGAAGGCGA TGATATGCCA 17520  
 40 AAGCAACAAG CACAACAAAC AGATATTGCT CAAAAAGCAA AAGATCTTTT CCGTGAAGAA 17580  
 ACTGTACATG TGATAGATGA AGAGTGATAC ATGACAAGCG ATATAATCGT ATGTATAATG 17640  
 AAAGAAACAT CATTTTATTG ATAAATATTT ATTGATTTTC AAGGAGGAAA TGGAATATGC 17700  
 45 GCGGTGGCGG AAACATGCAA CAAATGATGA AACAAATGCA AAAAATGCAA AAGAAAATGG 17760  
 CTCAAGAACA AGAAAACTT AAAGAAGAGC GTATTGTAGG AACAGCTGGC GGTGGCATGG 17820  
 TTGCAGTTAC TGTAACCTGT CATAAAGAAG TTGTCGACGT TGAATCAAA GAAGAAGCTG 17880  
 50 TAGACCCAGA CGATATTGAA ATGCTACAAG ACTTAGTGTT AGCAGCTACT AATGAAGCGA 17940

55

	TCCCTGGaAT GTGATCATAG ATGCATTATC CAGAACCTAT ATCAAAACTT ATTGATAGCT	18060
	TTATGAAATT GCCAGGCATT GGTCCAAAGA CAGCCCAACG TCTGGCTTTT CATACCTTAG	18120
5	ATATGAAAGA AGACGATGTT GTTCAGTTTG CCAAAGCATT AGTAGATGTT AAGAGAGAAT	18180
	TAACATATTG TAGCGTATGT GGTCAACATTA CTGAAAATGA TCCATGTTAT ATTTGTGAAG	18240
	ATAAGCAAAG AGATCGTTCA GTTATTTGTG TTGTGGAAGA TGACAAAGAT GTCATAGCTA	18300
10	TGGAAAAAAT GAGAGAATAC AAAGGTTTAT ATCACGTTTT ACATGGGTCT ATTTGCGCTA	18360
	TGGATGGCAT TGGACCAGAA GATATTAATA TTCCTTCATT GATTGAACGC TTGAAAAACG	18420
	ATGAAGTTAG CGAATTAATC TTAGCTATGA ACCCGAACTT AGAGGGGGAA TCTACAGCCA	18480
15	TGTATATTTT TAGATTAGTT AAGCCTATAG GTATCAAAGT GACGAGATTA GCACAAGGGT	18540
	TATCGGTAGG TGGCGATTTA GAGTATGCTG ACGAAGTAAC ATTATCTAAA GCAATCGCAG	18600
	GTAGAACAGA AATGTAATkT CTCTATTAA ACATTTTTGA TTTAATACT ATAGTAAGAA	18660
20	AAGTCACAGT GTAATCATTG TGGCTTTTTT TATGGTGTGG TGTGATGTAC TACTTTATTT	18720
	GCGGTGTGGC GGTGGTATGG TTTACCTAGT TTTACTGAGG GATGGGTAAT CTTTAGGAAG	18780
25	CAAGCCGTTG GTTGTGATTT GTTACTTCTA ATAGTAATGA TGTGAATTGG ATTATCGAAT	18840
	TAGATCTATG GTTATGGTGT GTTGGTGCTA TTAATTTGAT AAATGCGGTT AATGACTATG	18900
	CAAATGAAAT TCTTTTGTA TGAATGAT AGATGCTGGC TTAGTAAGTT GTACTTCTTT	18960
30	GGTCTAAAGC TTATTAAATC AGCCTGTATA GCGGTGTTTT GAGAGATTAT TTAAAACTTG	19020
	TAAATTTATT TTTAATTTCT GGTAAAAAAA TAACGTCTCG TTTTGCCTTT TTTTGTATTG	19080
	ATATGGTTAG AGAAAAATCT GTTTCTTGTT CTAAAAACG TACTATTTAT AAGTGGGGAT	19140
35	TTTTTAAGTT CGATTTTATG GATAAGGGCG TTCAGTACAG ATGACAAAGG TGTAATTTTT	19200
	ACTGTTGTTA AGCAGTTTGA AAGCCTGTAT AGTATTTATT TGTTGAGGCA AACAAAACAA	19260
	CTCAACTTAA GAAATAACTT GAATTACTAA CGAAAATTAA TTTTAAAAAG TTATTGACTT	19320
40	AAATGTTAAT AAAATGTATA ATTAATTCTT GTCGGTAAGA AAAATGAACA TTGAAAACTG	19380
	AATGACAATA TGTCACAGTT AATTCCAAAA AACGTAACTA TAAGTTACAA ACATTATTTA	19440
	GTATTTATGA GCTAATCAAA CATCATAATT TTTATGGAGA GTTTGATCCT GGCTCAGGAT	19500
45	GAACGCTGGC GGCGTGCTTA ATACATGCAA GTCGAGCGAA CGGACGAGAA GCTTGCTTCT	19560
	CTGATGTTAG CGGCGGACGG GTGAGTAACA CGTGGATAAC CTACCTATAA GACTGGGATA	19620
	ACTTCGGGAA ACCGKAGCTA ATACCGGATA ATATTTTGAA CCGCATGGTT CAAAAGTGAA	19680
50	AGACGGTCTT GCTGTCACTT ATAGATGGAT CCGCGCTGCA TTAGCTAGTT GGTAAGGTAA	19740

55

## EP 0 786 519 A2

	GAGACACGGT CCAGACTCCT ACGGGAGGCA GCAGTAGGGA ATCTTCCGCA ATGGGCGAAA	19860
	gCtGaCGGAG CAACGCCGCG TGAGTGATGA AGGTCTTCGG ATCGTAAAAC TCTGTTATTA	19920
5	GGGAAGAACA TATGTGTAAG TAACTGTGCA CATCTTGACG GTACCTAATC AGAAAGCCAC	19980
	GGCTAACTAC GTGCCAGCAG CCGCGGTAAT ACGTAGGTGG CAAGCGTTAT CCGGAATTAT	20040
	TGGGCGTAAA GCGCGCGTAG GCGGTTTTTT AAGTCTGATG TGAAAGCCCA CGGCTCAACC	20100
10	GTGGAGGGTC ATTGGAACCT GGAAAACTTG AGTGCAAGAG AGGAAACTGG AATTCCATGT	20160
	GTAGCGGTGA AATGCGCAGA GATATGGAGG AACACCAGTG GCGAAGGCGA CTTTCTGGTC	20220
	TGTAAGTGAC GCTGATGTGC GAAAGCGTGG GGATCAAACA GGATTAGATA CCCTGGTAGT	20280
15	CCACGCCGTA AACGATGAGT GCTAAGTGTT AGGGGGTTTC CGCCCCTTAG TGCTGCAGCT	20340
	AACGCATTAA GCACTCCGCC TGGGGAGTAC GACCGCAAGT TGAAACTCAA AGGAATTGAC	20400
	GGGGACCCGC ACAAGCGGTG GAGCATGTGG TTTAATTGCA AGCAACGCGA AGAACCTTAC	20460
20	CAAATCTTGA CATCCTTTGA CAACTCTAGA GATAGAGCCT TCCCCTTCGG GGGACAAAGT	20520
	GACAGGTGGT GCATGGTTGT CGTCAGCTCG TGTCGTGAGA TGTTGGGTGA AGTCCCGCAA	20580
25	CGAGCGCAAC CCTTAAGCTT AGTTGCCATC ATTAAGTTGG GCACTCTAAG TTGACTGCCG	20640
	GTGACAAACC GGAGGAAGGT GGGGATGACG TCAAATCATC ATGCCCTTA TGATTTGGGC	20700
	TACACACGTG CTACAATGGA CAATACAAAG GGCAGCGAAA CCGCGAGGTC AAGCAAATCC	20760
30	CATAAAGTTG TTCTCAGTTC GGATTGTAGT CTGCAACTCG ACTACATGAA GCTGGAATCG	20820
	CTAGTAATCG TAGATCAGCA TGCTACGGTG AATACGTTCC CGGGTCTTGT ACACACCGCC	20880
	CGTCACACCA CGAGAGTTTG TAACACCCGA AGCCGGTGGA GTAACCTTTT AGGAGCTAGC	20940
35	CGTCGAAGGT GGGACAAATG ATTGGGGTGA AGTCGTAACA AGGTAGCCGT ATCGGAAGGT	21000
	GCGGCTGGAT CACCTCCTTT CTAAGGATAT ATTCGGAACA TCTTCTTCAG AAGATGCGGA	21060
	ATAACGTGAC ATATTGTATT CAGTTTTGAA TGTTTATTTA ACATTCAAAT ATTTTTTGTT	21120
40	TAAAGTGATA TTGCTTATGA AAATAAAGCA GTATGCGAGC GCTTGACTAA AAAGAAATTG	21180
	TACATTGAAA ACTAGATAAG TAAGTAAAT ATAGATTTTA CCAAGCAAAA CCGAGTGAAT	21240
	AAAGAGTTTT AAATAAGCTT GAATTCATAA GAAATAATCG CTAGTGTTTCG AAAGAACACT	21300
45	CACAAGATTA ATAACGCGTT TAAATCTTTT TATAAAAGAA CGTAACTTCA TGTTAACGTT	21360
	TGACTTATAA AAATGGTGGA AACATAGATT AAGTTATTAA GGGCGCACGG TGGATGCCTT	21420
	GGCACTAGAA GCCGATGAAG GACGTTACTA ACGACGATAT GCTTTGGGGA GCTGTAAGTA	21480
50	AGCTTTGATC CAGAGATTTT CGAATGGGGA AACCAGCAT GAGTTATGTC ATGTTATCGA	21540

55

	GAGGAAGAGA AAGAAAATTC GATTCCCTTA GTAGCGGCGA GCGAAACGGG AAGAGCCCAA	21660
	ACCAACAAGC TTGCTTGTTG GGGTTGTAGG ACACTCTATA CGGAGTTACA AAGGACGACA	21720
5	TTAGACGAAT CATCTGGAAG GATGAATCAA AGAAGGTAAT AATCCTGTAG TCGAAAATGT	21780
	TGTCTCTCTT GAGTGGATCC TGAGTACGAC GGAGCACGTG AAATTCCGTC GGAATCTGGG	21840
	AGGACCATCT CCTAAGGCTA AATACTCTCT AGTGACCGAT AGTGAACCAG TACCGTGAGG	21900
10	GAAAGGTGAA AAGCACCCCG GAAGGGGAGT GAAATAGAAC CTGAAACCGT GTGCTTACAA	21960
	GTAGTCAGAG CCCGTTAATG GGTGATGGCG TGCCTTTTGT AGAATGAACC GGCGAGTTAC	22020
	GATTTGATGC AAGGTTAAGC AGTAAATGTG GAGCCGTAGC GAAAGCGAGT CTGAATAGGG	22080
15	CGTTTAGTAT TTGGTCGTAG ACCCGAAACC AGGTGATCTA CCCTTGGTCA GGTTGAAGTT	22140
	CAGGTAACAC TGAATGGAGG ACCGAACCGA CTTACGTTGA AAAGTGAGCG GATGAACTGA	22200
	GGGTAGCGGA GAAATTCCAA TCGAACCTGG AGATAGCTGG TTCTCTCCGA AATAGCTTTA	22260
20	GGGCTAGCCT CAAGTGATGA TTATTGGAGG TAGAGCACTG TTTGGACGAG GGGCCCCCTCT	22320
	CGGGTTACCG AATTCAGACA AACTCCGAAT GCCAATTAAT TTAACCTGGG AGTCAGAACA	22380
	TGGGTGATAA GGTCCGTGTT CGAAAGGGAA ACAGCCCAGA CCACCAGCTA AGGTCCCAAA	22440
25	ATATATGTTA AGTGGAAGAG GATGTGGCGT TGCCAGACA ACTAGGATGT TGGCTTAGAA	22500
	GCAGCCATCA TTTAAAGAGT GCGTAATAGC TCACTAGTCG AGTGACACTG CGCCGAAAAT	22560
30	GTACCGGGGC TAAACATATT ACCGAAGCTG TGGATTGTCC TTTGGaCAAT GGtAGGAGAG	22620
	CGTTCTAAGG GCGTTGAAGC ATGATCGTAA GGACATGTGG AGCGCTTAGA AGTGAGAATG	22680
	CCGGTGTGAG TAGCGAAAGA CGGGTGAGAA TCCCCGCCAC CGATTGACTA AGGTTTCCAG	22740
35	AGGAAGGCTC GTCCGCTCTG GGTTAGTCGG GTCCTAAGCT GAGGCCGACA GcGTAGGCGA	22800
	TGGATPAACAG GTTGATATTC CTGTACCACC TATAATCGTT TTAATCGATG GGGGGACGCA	22860
	tAGGATAGGC GAAGcGTGcG ATTGGATTGC ACGTCTAAGC AGTAAGGCTG AGTATTAGGC	22920
40	AAATCCGGTA CTCGTTAAGG CTGAGCTGTG ATGGGGAGAA GACATTGTGT CTTGAGTCG	22980
	TTGATTTTAC ACTGCCGAGA AAAGCCTCTA GATAGAAAAT AGGTGCCCCG ACCGCAAACC	23040
	GACACAGGTA GTCAAGATGA GAATTCTAAG GTGAGCGAGC GAACTCTCGT TAAGGAACTC	23100
45	GGCAAAATGA CCCCCTAAGT TCGGGAGAAG GGGTGCTCTT TAGGGTTAAC GCCCAGAAGA	23160
	GCCGCAGTGA ATAGGCCCAA GCGACTGTTT ATCAAAAACA CAGGTCTCTG CTAAACCGTA	23220
	AGGTGATGTA TagGGcTGAC GCCTGCCCCG TGCTGGAAGG TTAAGAGGAG TGGTTAGcTT	23280
50	CTGCGAAGcT ACGAATCGAA GCCCCAGTAA ACGGCGGCCG TAACTATAAC GGTCCTAAGG	23340

55



	TGTCTCAACG	AGAGACTCGG	TGAAATCATA	GTACCTGTGA	AGATGCAGGT	TACCCGCGAC	23460
	AGGACGGAAA	GACCCCGTGG	AGCTTTACTG	TAGCCTGATA	TTGAAATTCG	GCACAGCTTG	23520
5	TACAGGATAG	GTAGGAGCCT	TTGAAACGTG	AGCGCTAGCT	TACGTGGAGG	CGCTGGTGGG	23580
	ATACTACCCCT	AGCTGTGTTG	GCTTTCTAAC	CCGCACCACT	TATCGTGGTG	GGAGACAGTG	23640
	TCAGGCGGGC	AGTTTGACTG	GGGCGGTGCG	CTCCTAAAAG	GTAACGGAGG	CGCTCAAAGG	23700
10	TTCCCTCAGA	ATGGTTGGAA	ATCATTTCATA	GAGTGTAAG	GCATAAGGGA	GCTTGACTGC	23760
	GAGACCTACA	AGTCGAGCAG	GGTCGAAAAG	CGGACTTAGT	GATCCGGTGG	TTCCGCATGG	23820
	AAGGGCCATC	GCTCAACGGA	TAAAAGCTAC	CCCGGGGATA	ACAGGCTTAT	CTCCCCCAAG	23880
15	AGTTCACATC	GACGGGGAGG	TTTGGCACCT	CGATGTCGGC	TCATCGCATC	CTGGGGCTGT	23940
	AGTCGGTCCC	AAGGGTTGGg	CTGTTGCCCC	ATTAAAGCGG	TACGCGAGCT	GGGTTTCAGAA	24000
	CGTCGTGAGA	CAGTTCGGTC	CCTATCCGTC	GTGGGCGTAG	GAAATTTGAG	AGGAGCTGTC	24060
20	CTTAGTACGA	GAGGACCGGG	ATGGACATAC	CTCTGGTGTA	CCAGTTGTCT	TGCCAACGGC	24120
	ATAGCTGGGT	AGCTATGTGT	GGACGGGATA	AGTGCTGAAA	GCATCTAAGC	ATGAAGCCCC	24180
25	CCTCAAGATG	AGATTTCCCA	ACTTCGGTTA	TAAGATCCCT	CAAAGATGAT	GAGGTTAATA	24240
	GGTTCGAGGT	GGAAGCATGG	TGACATGTGG	AGCTGACGAA	TACTAATCGA	TCGAAGACTT	24300
	AATCAAAATA	AATGTTTTGC	GAAGCAAAT	CACTTTTACT	TACTATCTAG	TTTTGAATGT	24360
30	ATAAATTACA	TTCATATGTC	TGGTGACTAT	AGCAAGGAGG	TCACACCTGT	TCCCATGCCG	24420
	AACACAGAAG	TTAAGCTCCT	TAGCGTCGAT	GGTAGTcGAA	CTTACGTTCC	GCTAGAGTAG	24480
	AACGTTGCCA	GGCAAAAAAT	GGATGCGATG	AGCCGCATTG	AGACCGCAAG	GTCTCTTTTT	24540
35	TTTATGTCTA	AAACGTCAAA	ATAAAAAGCA	AACACAAAGA	AAAATGGCTT	GGCGAAGTGA	24600
	AAACGTTTGA	ATCTGACGAA	ACGAGAAAAG	ArcGCAACGA	GTTTAGTAGA	GCTAAATGAG	24660
	TAAGyGAGAG	CCGAAGrAGA	GGAAAGAAGC	AAGCGATTGT	CACAAGTCAA	GAAAGGTTCT	24720
40	TAGCGAsGAT	GGTAGCCAAC	TTACGTTCCG	CTAGAGTAGA	ACTGGAAATG	ATAATTTAAT	24780
	AATGTACACT	TTCGATTGTC	TAAGTATGTA	CAACTTTAAT	TTTGTGTTTA	TATAAAITTA	24840
	AAATGATATC	ATCGAAAACA	AAATATTGTA	TAAATAGAGA	AGAGCAGTAA	GACGGTATCT	24900
45	AATTGAAAAT	GATCTTACTG	CTCTTTTATA	TACTTTATTG	AAATACAAAA	AGGAAATTAA	24960
	TTATTATACA	ATAGACAAGC	TATTGCATAA	GTAACACTAA	CTTTTATCAA	AGAAGTGTTA	25020
	CTTTATAATT	AATGATTTTA	TTAGAGCGTC	TACATGCGGT	TTTAAAGCAT	CATCGTCTAT	25080
50	ACCGCCAAAG	CCTAATATAA	ATTTAGGGGT	TTTCTTATAG	TCTTGATCAT	CATCAAAATT	25140

55

	TCCATTTTTT ACTGTAATTG TAAAATGCAT ACCCGTTTCA GCACCTTGAA TATCAAGCTG	25260
	CTCTTTGTAA GGTTCGAATC TTTTAAAAT ATAGGTTAGT TTTCTACGAT AAATTCGTCT	25320
5	CATTTTATTT AAATGCCTTT CAAAACCACC GGAAGATATA AACGTTGCAA TAAGGTTTTG	25380
	CATATGAACA GGTACAGTGT TGCCTTCAAT GTGATTTTGA GAATGATATT TTTTCATTAT	25440
	AGAATAGGGT AACACCATAT ATGCAACTCG ACAGCTAGGA AAAATAGACT TTGAAAATGT	25500
10	ACTGATATAA ATCACTTTTT CTCCTCTTGA ATATAGACCT TGAATTGCTG GAATGGGTTT	25560
	GCCGAAATAT CTAAACTCGG AATCATAATC ATCTTCTATA ATAAATCGTT CTTCTTTTTT	25620
	TTGAGCCCAT TGTATTAATT GAGTTCGTTT TTTTAAGTCC ATCACATATC CAGTTGGAAA	25680
15	TTGATGGGAA GCGGTTATAT ATACTATATT TTTTGTGAT TTAATAACTT CATCTACGTT	25740
	TATTCATTA TCTTCAACTT CAATTGTTC ATATTCAACT TGTTTTTTAT CTAAAATATT	25800
	TTTGATTGGT GGATAACTAG GTTTTTCGAT AATAAATGTT GAAGTATAAA GTAAATCGAC	25860
20	TAATTGATTT ACTAATTGTT CGGTAGATGA GCCAATTATA ATTTGATTAG GATCACAAAT	25920
	TACGCCACGA TTAGTAAATA AATAAAATGC CAGTTGAAAC CGCAAATGTA ATTCTCCTTG	25980
25	AAAATGTCCT CTACGTAATT GATTAAATG ATTTGTATCA TAAAGATCTT TGAATACTT	26040
	TCTGAAAAGT TCTATAGGGA AATGTTTCGT ATCTATTCA TCCAAATTAA AAGCATAATC	26100
	ATAAGCTTCA TCACTCGCTT TTGGTTTATA TGAATCATCA TCAAAAAGAG AGGGGATAGG	26160
30	TTGATTGTTT AAAATTGTTA AAGATTCAAT TTCGGACACA AAATATCCAG AGCGAGGTCT	26220
	TGAATAAATG TAACCTTCGT CTAATAGAAG TTGATATGCA TGCTCTACGG TTGTTTGGCT	26280
	AATAGATAAA TGTTTGCTTA ATTGTCTTTT AGAATAAAAT TTATCGCCTT CTTTAAATTG	26340
35	ACCTTCAATT ATTTGTTTTT TTAATTTTTT ATAAAGTTGA TGGTATAAAG TGTTTTTCAA	26400
	TTTTATAACT GACCTCCTAA ATTTATCTTA TTTTGTACCT TTTTAAATAT CAGTTTATAC	26460
	ATTACAATGT ATTTAATCAA CTTGAAAAGG GGTTTTATGT ATAATGAGTA AAATTATTGG	26520
40	ATCAGACAGA GTCAAAAGAG GTATGGCTGA AATGCAAAAA GCGGCGGTTA TTATGGATGT	26580
	CGTTAATGCT GAGCAAGCAA GAATTGCAGA AGAAGCTGGC GCGGTAgCAG TTATGGCATT	26640
	AGAACGAGTA CCTTCTGATA TTAGAGCTGC TGGTGGTGTT GCACGTATGG CAAACCCTAA	26700
45	AATTGTAGAA GAAGTAATGA ATGCTGTTTC TATTCCAGTC ATGGCTAAAG CACGTATTGG	26760
	TCATATCACT GAAGCAAGAG TATTAGAGGC GATGGGTGTT GACTATATTG ATGAATCAGA	26820
	AGTGTTAACA CCAGCAGATG AGGAATATCA CTTAAGAAAA GATCAATTTA CAGTACCATT	26880
50	TGTATGTGGA TGTCGTAATT TAGGTGAAGm TGCGCGTAGA ATTGGTGAAG GTGCTGCTAT	26940

55

	ACAAGTTAAT TCAGAAGTTA GTCGATTGAC TGTAATGAAT GATGATGAGA TTATGACTTT	27060
	TGCGAAAGAT ATCGGTGCGC CTTATGAAAT TTTAAACAA ATTAAAGACA ATGGTCGTTT	27120
5	ACCGGTAGTT AACTTTGCAG CTGGTGGCGT TGCGACTCCT CAAGATGCTG CTTTAATGAT	27180
	SGAATTAGGT GCTGACGGTG TATTCGTTGG ATCAGGTATT TTTAAATCAG AAGATCCAGA	27240
	AAAAATTGCT AAAGCAATTG TTCAAGCAAC AACACATTAC CAAGACTATG AACTAATTGG	27300
10	AAGATTAGCA AGTGAACCTG GCACTGCTAT GAAAGGTTTA GATATCAATC AATTATCATT	27360
	AGAAGAACGT ATGCAAGAGC GTGGTTGGTA AGATATGAAA ATAGGTGTAT TAGCATTACA	27420
	AGGTGCAGTA CGTGAACATA TTAGACATAT TGAATTAAGT GGTCAATGAAG GTATTGCAGT	27480
15	TAAAAAAGTT GAACAATTAG AAGAAATCGA GGGCTTAATA TTACCTGGTG GCGAGTCTAC	27540
	AACGTTACGT CGATTAATGA ATTTATATGG ATTTAAAGAG GCTTTACAAA ATTCAACTTT	27600
20	ACCTATGTTT GGTACATGCG CAGGATTAAT AGTTCTAGCG CAAGATATAG TTGGTGAAGA	27660
	AGGATACCTT AACAAAGTTGA ATATTACTGT ACAACGAAAC TCATTCCGTA GACAAGTTGA	27720
	CAGCTTTGAA ACAGAATTAG ATATTAAAGG TATCGCTACA GATATTGAAG GTGTCTTTAT	27780
25	AAGAGCCCCA CATATTGAAA AAGTAGGTCA AGGCGTAGAT ATCCTATGTA AGGTTAATGA	27840
	GAAAATTGTA GCTGTTTACG AAGGTAAATA TTTAGGCGTA TCATTCCATC CTGAATTAAAC	27900
	AGATGACTAT AGAGTAACTG ATTACTTTAT TAATCATATT GTAAAaAAAG CATAGCTTAA	27960
30	TGTATGCTAA ATCAACGAAT TATTGATATT TATAGATTTG TTGAGAAGAA AATATCTCCT	28020
	TCAAACCTAG CTTTGGAGGA GTTATTTTTT ATGTCAAAAT TAAAAATGAT AAAAAATAAA	28080
	GCTATACATA AGAAAAAAAC CCTTCAAAGA GACTGAGAAT AGTCAAAATT TTGAAGGGGT	28140
35	TAATTCGATG TTGATGTATT TGTAAATAA AGAATCcAGC GATTGCAGCT GAAATGAAAG	28200
	ATACTAGTGT tGCACCGAAT AATAATTCA AACCAAAGCG GGCAACTGTA TCTCCTTTTT	28260
40	TGTCATTAAG TGATTTAATC GCACCTGAAA TAATACCGAT AGAGCTAAAG TTAGCAAATG	28320
	ATACTAAGAA TACAGATGTA ACACCTTTTG CGTGTCAGA TAAATCACTA AGTTTACCAA	28380
	GTGCTTGCAT TGCTACAAAT TCGTTAGATA ATAGTTTTGT CGCCATAACT GAACCGGCTT	28440
45	GAACTGCATC TTGCCATGGC ACACCGACTA AGAATGCAAA TGGTGCAAAG ACAAACCAA	28500
	TTAATGTTTG GAAATCCCAA GAAATAGCGC CACCTGAAAC TGTACTAAAG ATATTGCTTA	28560
	CAATTCCATT TAATAGAGCG ATAATGGCAA TGTATCCGAT TAACATTGCG CCTACAATGA	28620
50	CAGCTACTTT AAATCCATCT AAAATATATT CTCCTAGCAT TTCGAAGAAT GATTGTTGTC	28680
	TTTCTTCAGT TTCTTCAACT AATAATTTGT CATCTTCTTC ATTAACTTTA TAAGGGTTAA	28740
55		

TAGGTTCAAT TAAGGTAAAG TATGCACCGA TAATTGAAGC AGAAACAGTC GACATTGCTG 28860  
 AAGCTGTTAA TGTGTATAAA CGTTGCTTAG GTATGTATGG TAATTGTTTT TTAATTGAAA 28920  
 5 TAAATACTTC AGATTGTCCC AAAATTGCTG CAGCAACTGC ATTGTATGAT TCTAAACGTC 28980  
 CCATACCATT AATTTTAGAA ATTAAGAATC CTAAACATT AATGATTAAA GGTAATCT 29040  
 TTGTGTATTG AAGGATACCG ATAATCGCTG AAATAAATAC GATAGGTAAT AATACACTGA 29100  
 10 AGAAGAATGG TGGTTGCTTA GGATCGATAT ATTGAATACC ACCGAATACA AAGTTAACAC 29160  
 CATCTGCTGC TTTTAATAAT AAGTAGTTAA AACCGTTTGA AATACCACCA ATAACCTTGA 29220  
 TTCCCATTTGT AGTTTTAAGC AAGATAAATG CAAAGATAAG CTGAATTGCA AGTAAATTC 29280  
 15 CTACATATTT CCAGCGAATA TTTTTCCTGT CTGAGCTAAA TAGAAACGCA AGTGCTAAAA 29340  
 AGAAGATAAT TCCGATAATC CCAATTAGAA TATGCATATA TTTCTCATTC CTTTAGTTTT 29400  
 TTCTACaATc TATCATACAA TAAAATGGAA GGGCTAACAT CATAAATTTT TGAAAATATA 29460  
 AAAACAAATT AATTGAAAAA GGTCAAAATA GGTCAATATA TATAGTCAAA GAAGGTCAAA 29520  
 AAGGGGTGAT ATACATGCAC AATATGTCTG ACATCATAGA ACAATAaTCA AACGTTTATT 29580  
 25 TGAAGAGTCG AATGAAGATG TCGTTGAAAT TCAGAGAGCG AATATCGCAC AGCGTTTTGA 29640  
 TTGCGTACCA TCACAATTAA ATTATGTAAT CAAAACACGA TTCACTAATG AACATGGTTA 29700  
 TGAAATCGAA AGTAAACGTG GTGGTGGTGG TTACATCCGA ATCACTAAAA TTGAAAATAA 29760  
 30 AGATGCAACA GGTATATTA ATCATTTGCT TCAGCTGATT GGACCTTCTA TTTCTCAACA 29820  
 ACAAGCTTAT TATATTATTG ATGGGCTTTT AGATAAAATG TTAATAAATG AACGTGAAGC 29880  
 TAAAATGATT CAAGCAGTTA TTGATAGAGA AACGCTATCA ATGGATATGG TTTCTAGAGA 29940  
 35 TATTATTAGA GCAAATATTT TAAAACGTTT GTTACCAGTT ATAAATTATT ACTAAATGAA 30000  
 ATGAGGTGTT GAAGTGCTTT GTGAAAATTG TCAACTTAAT GAAGCGGAAT TAAAAGTTAA 30060  
 AGTTACAAGT AAAAATAAAA CAGAAGAAAA AATGGTGTGT CAAACTTGTG CTGAGGGGCA 30120  
 CCATCCGTGG AATCAAGCTA ATGAACAACC TGAaTATCAA GAACATCAAG ATAATTTCGA 30180  
 AGAAGCATTT GTTGTTAAGC AAATTTTACA ACATTTAGCT ACGAAACATG GAATTAATTT 30240  
 45 TCAAGA 30246

## (2) INFORMATION FOR SEQ ID NO: 57:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14333 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

	TATTCCTCCCA TCGGTTTATT AAATCGTCCA TTTCAATACT GTTTTTCCTCC AAGATGTCGA	60
5	TAAATCCATT TCAAACGCTT GGACGATATC TTGCATCGTA CATAATTAA TTTCATGTCC	120
	TTTAAATAAT GCTAACTTTT CAACTATGTC TGGGTACTTA CGATATAAAT CAACAACTTG	180
10	CTCAAAATCT TTAGAGCCGC TTCGACTACT ACCAATCAAC GTTAATCCTT TTCAAGTAC	240
	TAATCGTGTA TTCACTTCCA CGGGTAATTC ACTTACGCCT AACAAAGCAA TACTGCCTTC	300
	TGGTGAAATA TGTTCACATA TTTGTTGAAG TGCAACTTGA CTTCTTTTAC CTCCAACACA	360
15	TTCAAAATGCA TGATCAATTT TAAGATCATC TGGTATTTGA TTTACTGTAA AGATGTCATC	420
	TACAAATGAA AAATGACTTA ATTTATAGTC TGTCTTACCA AATACATAAG TTTTAGCTTC	480
	TGGGTACAAC TTACGTAGCA AAATAGCAGT AATATAACCT AAGTTACCAT CACCCCAAAT	540
20	ACCAAAGCTG GTTTTCAAAG GTATAGATTT ACGTTCAAAT CGTTGTATAG CATGATAACT	600
	TACTGACACT AACTCTGTGT ATGAAATCGT ACTCAAATCA ATGTCATTAG GCAGCGGAAC	660
	GATACGATCA TGTGCCATCA CAACGTAGTC TTGCATAAAA CCATCATAAC CACTAGATCT	720
25	AAAATAACTA GAGGCTAAGT AATTCCTCCG AATAATATGA TGTGCTCTG TAGGTGTATT	780
	CGGTACCATT ACTACTTTCG TACCTTTTTC AAATACCCCT TTAATATCAA ATACAACCTC	840
	ACCAACAGCT TCATGAACTA ATGACATTGG TAATTTTTTG CGTAGTACAT TTTCATCTCT	900
30	TCGACCTGTG TAATACCTTT GATCAGCTGC ACAAATAGAC AAGTATAAAG GTCTTACGAT	960
	GACATGATTA CCATAAATAT CAACATTATT ATATGTGACG TCGAACTGTC TCGGTGCAAC	1020
35	GAGTTGATAT ACTTGATTAA TCATCGGCAA TATCACCTTG AATAATGGCA TTTGCTACTT	1080
	TTAAATCATA CGGTGTTGTC ACTTTAATGT TGTATAGTTC TCCaCGTACC AATTTAACTG	1140
	CATGTCCAGA TTCGACAATG ATTTTACATG CATCTGATAA GATTTCTTTT TGTTCACTAC	1200
40	TTAAGGCGCG ATAACATCTT TGTAAATAAT TAATATTAAA TGATTGTGGT GTTTGGCCTT	1260
	GATACATTTT ATTCCTTACA GGGATACTGT GTATGTTCTG TTTATCTTTA GACATTACAA	1320
	TCGTATCAAT TGCTTCAATG ACTGTATCTA CTGCACCATA TTTTGCTGCT ACTTCAATGT	1380
45	TCTCTTTAAT AATACGTTGA GTTAAAAATG GTCTTACGGC ATCATGAGTT ACAATCACAT	1440
	CATCATTATT AATTCATTT ACATTGCGAA TATGGTCGAT AATGTTTATA ATTGTTTCAT	1500
	TTGATCCGT ACCACCTGCA ACTACTTTGA CACGTTGATC TGTAATGTTA TATTTTTTTA	1560
50	AAATATCCTG TGTATGGGAA ATCCACTGTG CTGGCGTTGC GATAATAATC TCATTAAATT	1620
	CACTCACTAA AATGAACTTC TCAATTGTAT GGATTAAAA CGGTTTATTA TCAATATCTA	1680

55

	CTGCATAAAT CATGTTGTCC TCCATTCTGT CATTACATCA TTTCCATTTA TACATTACTG	1800
	ACCTATGCCC GCACATAAGC CTAACCTATT GCTCACTTGC CTCTTTTATT AATCCAAAGA	1860
5	TAGTTGTCAC AATAGTGTGA TAAATTTTAA TAAAAATGTA TTTTGTAAAC TGACCATTCT	1920
	AAGTTGTTTT GCCATGCAGT TAATCATTAA CTCTGACGAT ATTAAATTGT TAAAGGTATT	1980
	AATGTTTACT CTTTTTCAAA TTCATTATTA CTGCCATCAT TTTACCATAT ATTATAATAA	2040
10	ATTTATCTTA TTAAGTGGCT GTACTTGATT TTCACTTTAA AAATTATCAA ATATTGCCAT	2100
	CTCATTTTAA GTATACAAAA TGCAAAACAA CCGATTCAAC AGCATATTTC ACACAAGTAA	2160
15	ACCGGCTATT TATCAACGTA TATTCGAAGA TGAATTATTT CGATAGTATC TATAGACCAG	2220
	ACGGCATTCT CACTTTCATA GCTATAACTA TACCAGCGTT TCGTCCTCA AAGGTGCATA	2280
	CTAATAAATC GTAAACATGA CTTTATCAAA TCGTTCCTTC TTGTTAACTA ATTTATCAAA	2340
20	TGTCTCCGGG CCTTTTCTA ACGGTAAAAA ATGAGAAATA ATAGGCTTTA CATTAAATATC	2400
	TTTCGTCTTC ATATAATGTA AGGTTGCCGT CCACTCTTTG CCCGGAAAAT TACTGGACAA	2460
	ACAGTTCCAA GAGCCACATA CTGTCAACTC GTTACGCAGA ATTTTTTCAA AATGAACGCG	2520
25	ATCAATCTCA ATATCATCAT ATGGTATTCC GAGTAATACC ACCTCGCCAC CTTTTTTAGG	2580
	TAGCGTCAAT ATTTGACCAA TCGTAACTTT AGCACCTGAT GATTCTATAG CTAAATCGAT	2640
	TTGATTGGCG TAATGATTTT CGATGAATTT CTCAAGATTT TCTTCTTTTG AATTGATTGT	2700
30	TTGATGTGCG CCCAATGATG TTGCAATATC TAGTTTATGC GCATCTATAT CTATAGCGAT	2760
	GATATGTGCA GCACCAAATA TTCGTGCCCA TTGAATAGCT AACAAACCTA TACTGCCACA	2820
35	CCCCATTACT GCAACAGTCA TACCAGGTTG TATATTGAT TTATAAAACC CATGCGCAAC	2880
	AACGGCTGAT GGCTCAACCA TTGCTGCTTC AATGTAATCA ACATTGTCTG GAACCTTTAA	2940
	AACATTTTGC GCTGGCAATT TGACATATTC CGCGAACGAT CCAGGTTTAT ATGAGCCAAT	3000
40	GACGAATAAC TTTTCACATC GTGCATATTC ACCTTTTAAA CAATACTCGC ATTGATAACA	3060
	AGGTATTGCT GGGCAACCTG TCACTTTGTC GCCCACATTA ACATGCGTAA CATCACTTCC	3120
	AATGGCATCT ACTACACCTG AAAATTTCATG ACCAAATGGC ATACCTTTAA TGTATGGCCC	3180
45	CATTTTTTTG TATCGTGACG TGTCTGAACC ACATATGCCA GTCGCTCGTA CTTTAATAAT	3240
	AACGTCATTC GCACTTTCAA TGAAGTGGCT TTCATTATCC TCATACCGTA AATCTTCCAC	3300
	GCCATATAAT TTCAATGCTT TCACTTGTA ATCACCTCAA ATTTGATTTA ATTCACAACT	3360
50	TTTTTCTTTT TAAAAATACC TGTGCAAAAA TAACCTGCAA TGACAATGGA ATTACTTACG	3420
	AGTAAATGTT CCAATAAAAA ATCAGTGATT TGTCTTAATG GCCCAAGCAT AAAAGTTAGC	3480
55		

	TGCTTTAATA CCTTCGCCGG ATTTTAAATG TTGATACGCC TCGTCCCATT TCGAAATATC	3600
	ATATATTTTT GTACACAAAG CTTACAGCATT TACTAAACCA TCCGCCATAA GTTGCAATGA	3660
5	AGGTTCCCAA TCTGCTGGCT TTTGACTTCT ACTACCAACA ACTGTTATTT CTTTTGAAT	3720
	CACTTTTTCC ATATCAAATG GAATTCAGC ATCCTTAAAA ATACCTATTT GACTGTAGAA	3780
	ACCTTTTTTG CGTAAATAT CCAAACCTTG TCGTGCTGCT GGAAGTGCAC CTGAACATTC	3840
10	AACAACAACA TCTGCACCGT AACCGTCTGT AATTCCATTG ATATACGTTT TTAAGTCTGT	3900
	TTGTTGTAAG TTAGTACAT AATCCATGTG CAATGCTTCT GCTTTATCTA ATCTGACTTT	3960
	GTCATTGTCC AATCCAGTTA CCACAACAGT TCGCCTTTA CTTTTTAACA CTTGTGCTAC	4020
15	AAGTAATCCG ATTGGCCCAG GTCCCATTAC AACTGCTACA TCGCCTGAAT TGAAGTGAAT	4080
	CTTAGAAACG CCATGATGTG CACATGCTAA TGGTCTGTG ATAGCTGCAG ACTGATACGA	4140
20	TATTCGCTCTG GAATATGATG CAACTTTTCT TCACGTGCAA TGACATAATT AGTAAATGCG	4200
	CCATCAACTT GTGTTCCAAT ACCTTTTCGA TGGTTGCATA AATTATAGTC TTTTGATTGA	4260
	CAGTATTCAC ACTCATTACA AACATAGAAT GTCGTTTCAG ATGTGACACG GTCACCAACT	4320
25	TTAAATCTT TAACGTCTGC TCCAACCTCA ACGATTCAC CAGAAAATTC ATGACCTAAT	4380
	GTCAGTGGA AATTAACTTT ATAATGACCT TCATAAGTAT GAATATCTGT GCCACAAATT	4440
	CCTGCATAAT GTACTTTAAT CTTTACTTTA TCATCTAGCG GTGTGCAAC TTCTTTATCA	4500
30	AGAAGTTCTA AGTTGCCATG TCCTTCTCTT GTTTTACTA AAGCTTTCAC CACAAACACC	4560
	TCGATTTTTA ATTGAATAGA CTAAATAGTT TAAAGATAAG ATAGTTAAG ATATTACCAC	4620
	CTTGATCAAT ACTTGAAATT TCAGATGAAC CTTTGGCAT TTGTACATTC GTACCTTTTCG	4680
35	CCATATCTGT GAAATGGGT GCTACGCTG TTGCAATATA TAGTGAAATT GCAATCATAA	4740
	TCGTACCCAC AATGACAGAA TGAATAATGT TTCCTCTGCG TGCACCAACA ATAAACGCGA	4800
40	CAACAAATGG TATCGTTGCT AAGTCACCAA AAGGTAGTAC TTGGTTTCCT GTTAAATATA	4860
	CGGCTAATAA AACAGTGATA GGTACTAAAA TTAATGCTGT CGAAATAACT GCTGGATGAC	4920
	CTAATGCTAC AGCCGCATCC AATCCAATAT AAATTCACG TTCGCCAAA CGTTTATTTA	4980
45	GCCATGTTCT TGCAGACTCT GAACTGGCA TTAAACCTTC CATTAGATT TTTACCATT	5040
	TAGGCATTAA TACCATTACT GCAGCCATTG ACATTCCTAA ATTAATGATG TCTCCAGGTT	5100
	TGTAACCTGC TAACACACCA ATACCTAAAC CTAAATTA GCGACAAAT ATAGACTCTC	5160
50	CAAATGCGCC AAAACGTTTT TGAATTGTTT CAGGATCAGC ATCTAACTTA TTCAGACCGG	5220
	GTACTTTTTG TAACAATTTA ACTAAGTAAA TACCTGGTGC ATAAGAAATT GTACTTCCTG	5280

55

	CTACTTTCAA ACAGATAATT TGGAAAATAA CTGCTGCTAA TAACGCTTGC CAAATACTGC	5400
	CTGATACGGC ATAAACCATT GCTGCTGTAA ACGTATAATG CCAAAAATTC CAAATATCTA	5460
5	CATTTCATCGT CTTTGTCACT TTAGTTACTA GCAATACAAC GTTAACTATG ATTCCGAGTG	5520
	GAATAATAAA TGCTGCGACA GATGATGCCC AAGCGATAGA TGATGTTGCT GGCCAACCTA	5580
10	CATCAATCAC ATTCAGACTG ACGCCTAAAT TTTTAACCAT CGCTTGTGCT GCTGGCCCTA	5640
	AATTTTAAAC TAATAAATCG ATGACTAAGA AAATCCCTAC AAAAGCCACA CCTATTGTTA	5700
	AACCAGACCT AAATGCCGCT CCAATTTTCT GCCTAAAGAA TAGGCCAAGC AAGAATATGA	5760
15	CAACCGGTAA AATAACAGT GCACCTAAAT CTAAAAATCC CCTTACAAA TCAGTGAAGT	5820
	AACTCATATT TAAACCCTCC CTGTTATATA TGCATTGTCA CGATACTTTC CGATTGTGAT	5880
	TACATTTGAC GTTACAGTCA TTTCAACGAC AACCCTTGCT AAATTCGACT GCAGTCCTTT	5940
20	TGAATTACAG TCACTGCGTT TCTATGTCAT CAACAATCAT TTGTCGTGAT AGTCATTTAT	6000
	ATGCAATTTG CATATATTAA TATGTTATCG ACCCACGTTA CATATCAATT CCGTTATTTT	6060
	TGTAACCTCTG TTAAGATTTG TTGTTTTGTT TCTTCAATAC CAATACCAGT TAAGAAATTA	6120
25	CGTGCGTTGA TAACTGGGAA TTTATATTCT TTTTGTGCA TTGCAGTTGT AACTAATAAA	6180
	TCTGCAGTGT CTTCATAAGG TCCAATTCT GTAATTTTGA TTTGTTTAAAT ATCTACTTTA	6240
	ATATTGTGTT CCTTTGCCAT TTCTTCAATT GCATTATTTA CTACTGTTGA CGTTGCAATA	6300
30	CCTGCACCAC ACGCTACTAA TACTTGTTC ATTTTCAATT CCTCCAATTA ATTTTITAGTT	6360
	ATATTCCAAA TAATCATTGA TTAGTGTGTC TAAAATTGTT TCATCTTTCG TTCGTAGAAT	6420
35	CTGCTCCAAT TTTTCTTCAC TTTGAAAAAT TTGCATCAAC TGTGTAACA GCTTAAGTTG	6480
	ATCATCTACT TTATCCATTG CTAACATAAA AACGATTTTC ACTTCTGTCT GTTGATCAAG	6540
	TGTTCCCAT TCAATAAACG GCACTTCTTT TTCTAGAACA GCCACACCTA TCGTTCTATG	6600
40	GTTAATATGT TCGACATCTG TATGCGGTAT AGCGACCGAA CATAGATGCG TTGGTAAACC	6660
	AGTAGCAAAT TCTTTTTCTC TGTCGATGAC TGCATCTTTA AACGTTGACT TCACGAACCC	6720
	ATTTTGAAAT AACACATCTG ACAATTTGTGA CAATACGGAT TCTTTATCAG TTGCCGACAA	6780
45	ATTGAGCATT ATATTTTCTT TATGCACTAA TTGCTGTCCC ATCCATTTTC CCTCGCTTCT	6840
	TTATTTGAAT AATTTTTTAA AATCTCATTT ACATCAGAAT TTTTGCGACT TTGTATGATG	6900
	CGCTTAATTG CGTCATTGTC TTGCGCCACA TCTCTCAATT GTAGTAACGC TCTTAAGTGT	6960
50	GTCACCTTAT CAACAGCAGC AATAGGTACA ATAATATGGA TTGCTGTGCC ATCTGACATG	7020
	TATATTGGTT CTTGTAATAT CAACATACTC ATCGCTGTTT TATGTACATG CTTTTTCAGAG	7080
55		



	TGCATCTCAT GAATATATTT AATATCAATA AAATGATTAG CAACTAACAC ATCACTTGCT	7200
	TTAGCAATAG CTTTCATCAAT ATTTTCAACA TGATGCATT CTTTCACGTG CCTTGCCGGT	7260
5	ATCAAGTCAG CTAAATCTAA TGyCTWATTT tGTGtGACaA TCGATCCATT AATGGTTGAA	7320
	ATTGAATTAT AATTGGCAAT AAAATCTTCT AAACCATCAC GTAGTcTGTA ATGTCATTAA	7380
10	CTGTGTTGT GCGTTCAATT AATGCCATTA ACTTGTtTAT TTCCTTATCA ATGTCAGCCG	7440
	ATTCTTtATt AATGTACTTC ATCACTTCTT TACGTAACtT TCGTTGCTCA TTTTCAGATA	7500
	AAGCTACTTT TGTGATAAAT AATTTTTTAT GTGTTAGGAC AAACATTGGT GAAAAGACGA	7560
15	TGTCATAATC TAATGTGTAA TTTTCAAATG TTCTAAGTGA AATCGCATCT AAGAAAATAA	7620
	TTTCTGGAAA TAAGTTTCGC AACTCGTATA ACATCATTTG TGATACTGAC GTGCCTTGtG	7680
	TACACACGAT AATAGCTTTT ATCTTGCCAT CGAAGTTTTC ATCTTGACGT CTCAAACtAC	7740
20	CTCCGAACAA CATGGTTAAA TATGCTATTT CATTATCAGG CAACGATTTT CCGAAATATT	7800
	CAGTTAACGA TTGACATGAT TGTTTCACCA TATGAAATAA GGATTGATAA TTTCTTGTA	7860
	AAGGATTTAT TAATTCATCA CGATCCGTTA AGTTATATTT AATCCTATAA AAAGCAGGCG	7920
25	TTAAATGTAA CAAGAGTTGC TGTGATAATT TCTCCTTATC TTCAATGTTA ATAAAAGTGA	7980
	TTTGTTCAAA ATGGTGAATC ATTTGAGCGA TGGCCATCGT TAAATTCGAT ATGCTATCTG	8040
	ATTCTTGCAA ATCAGTCCAT TGCACACTTG TTGAAAGTAA GTGTAATGTC AAATATAACT	8100
30	TTTCCGCTTC TGGCAAATCC GGCTCATGTT GCGTCATAAT CTCCGTGCT TGATATTCTT	8160
	TCGTATCCCT CAAATACTGA TAATTAATAT TTAATGGATT CATCACATGA CCACTTTGAA	8220
35	TTCGTCTACG AATCACACAA AGGACATAAG GCAATGAACT AAGTGATTTG TCTATAAAGC	8280
	GACTCTTCAA AAATTGTTCT ACCTGTTTGA TCTTGCTTTT TTGATATGCG ATATCTTCGA	8340
	ATGTTAAGTT GAGCGCCTTT AAAACTTCAC TTTTAGTAAT ATCATGATTC AACCTTTGAT	8400
40	CAATCAACTT AATGAAGAAA CGGCGAACTT CAAATTCATC ACCAACAATT TCATAACCAT	8460
	GTTTTCGAGA ATACTTAAGT GACAAACCAT GATTTTCCAA TTGCTCTTTC ACATGATTTA	8520
	TATCGTGAAT GACAGTATTT TTA CTGACTT GTAAATCAAT TGAAAAATGG TTTAGAGACA	8580
45	TTGCGTTTTT CTTACTAAAA AGCATGAGCA TTAAATAATA ACGACGTGTT TCTATGCTAA	8640
	AAATGACATT GTTGCCGTTT AACATTTGCT GCTCCGATAC ATCTCGCTTG AATAACGTCA	8700
	TGATTTcAGA ACTTACAATA AAATTTCTT GGCTTGTCTT TTCAAGTTTT GGATAACCCT	8760
50	CTTGTTCAAG CCACAAATTG ATTTTTTGAA TGCGATATCC TAGTTGTCTA CGAGACAAAC	8820
	CAAATATCGA TTCAAGTTCT TTACCATGAA TAGTAGGATT CAATACAATT TCTCTGAGTA	8880

55

	TCAATCGTCA CACCGATGTA CACACTTTGA ACACATATTT TCAAAATGAG CATGTACATC	9000
	ATTGTGATGT TTTAACAACA TTTCAATTAT ATCTATATTT TTTGTGATTT TAATCTTTTA	9060
5	AAATAAAGCA ATTGAAATTT TTGCATATAT TTTTGTGTTT TGTGTTTTTT TGAAGCATTT	9120
	TTAACATACA TATCTCAATC ATTATCAAAT TGTCTGACC ATTGTAACCC AATACAAAAA	9180
10	CCCTAAGGAC GCTTATATCA GCGCCTTAG GGTTAACTGT ATCTATTAA TTAAGTATTA	9240
	TTATTCGTAT GTACGTAAC TATGGTCTAT CAAGTTCCAC ACTTCTTCAA CATCAACTGC	9300
	TGTAGCAAAA TAAGCATTGG CAGGCTTACC TGTAACATGA TTTAAATCGA CAGCCATAGT	9360
15	GCCATAAGTT AGTGGACTTT GATGTTCAAT GTCGATATTA ACGGGTACCA TTGTAAACAA	9420
	TTCTGGTTGT AACAAATACA AAATTGTACA AGCATCATGT ATTGGACCAC CATCCATATT	9480
	AAAGTGAGTC TTGTATGTCT TCTTAAAGAA TTGCAATAAT TCTACGACGA ACTGTGCAAC	9540
20	AGGATTATTG ATACTTTCAA AGCGTTCAAT CACGTGATCG TCGGCTAAAA CTGATGTGT	9600
	TACATCTAAA CCAAACACAT TTATAGTAAT CCCACTTTCA AAAACACGCT TCGCTGCTTC	9660
	AGCATCTACC CAAATATTGA ATTCTGCTGT AGGCGTCCAA TTTCCAAATG TACCACCACC	9720
25	CATCAAAGTA ATAGATTCAA TATGCTCAGC GATTCTTGGC TCACGAATCA ATGCCGTTGC	9780
	TACATTGTA AGAGGACCTG TCGCTACAAT TGTTACAGGT GTATCACTCG TCATCACTTT	9840
	GTTTATAATC ACATCTGATG CTGGCATTGC AACTGCTTGA CGTGATGGTG TCGACGGTAG	9900
30	TTTCGGACCA TCTAATCCAG ATTCCCCATG TATTTAGAA GCAAAGGCAG CTGGTTTAAT	9960
	TAACGGCCTA TCCGCACCTT TCGCTACTGC TATATCTTGG CGTCCCATAA TATCCAATAC	10020
35	GTTCAAGGCG TTTGTGCTAT TCTTGTCAAC TGATTGATTA CCTGCGACTG TTGTTACAGC	10080
	TAATATCTCT AGTGGACTGT CAATTGCCCC CGCTAAAATT AATGCTATTG CATCATCGTG	10140
	TCCTGGATCA CAATCCATAA TAATCTTTCT TTTCACTTAT ATATCCACCT TTCTTAAGTT	10200
40	GTTATCGATA GCTTATGTAT ATTTATTTAT GTGGTGAATC ATGTTTATTT TGAAAAATAG	10260
	TTTTAACTTT CTCATATTTT TGGATACAAA CACTATTTAT CTATTTTATG GCTTATAAAT	10320
	TTATCCGATA TGCCTTATCA ACCTACCTCG CTAAAAATAG GATGTCTACA TATCTATACC	10380
45	CACTTTTGTC AACTCATTTT CACAACAATA TAAACAGCAA TTTATATGAT TGTTACATGA	10440
	TTCAAACAAT TTTTATGAAA AATATTTTCA TACACAGAAT ATATATTGAT ATTAAATTTT	10500
	TCAAAAGCTA TATTGAGAAT AATTAGGAGG GATGTTGATG AAATCTTTAT TTGAAAAAGC	10560
50	ACAGCAGTTC GGCAAGTCCT TTATGTTACC TATCGCAATC TTACCAGCTG CAGGTCTATT	10620
	GTTGGGTATC GGTGGTGCAT TAAGTAATCC AAACACCGTT AAAGCATACC CTATTTTAGA	10680

55

	AAATTTACCG	GTCTCTTTG	CAATTGGTGT	CGCAATCGGA	TTATCTAGAA	GCGATAAAGG	10800
	TACTGCAGGT	tTAGctGCGC	TGCTCGGTTT	CTTAATTATG	AACGCAACTA	TGAATGGCTT	10860
5	ATTAACATATC	ACGGGCACAT	TGGCAAAAGA	TCAGCTTGCA	CAAAATGGAC	AAGGCATGGT	10920
	GCTCGGTATA	CAAACGGTTG	AAACCGGTGT	TTTTGGCGGG	ATTATCACAG	GTATTATGAC	10980
	CGCAATACTT	CACAACAAAT	ATCACAAGT	GGTATTACCA	CCGTATTTAG	GTTTCTTTGG	11040
10	TGGCTCTAGA	TTTGTCCCTA	TTGTCACAGC	ATTGCGCGCA	ATCTTTTATG	GTGTATTGAT	11100
	GTTTTTCATT	TGGCCAAGCA	TACAAGCCGG	CATTTATCAT	GTTGGTGGAT	TTGTAACGAA	11160
15	AACAGGTGCC	ATCGGTACTT	TTGTTTATGG	CTTCATCTTA	AGATTGTTAG	GTCCACTCGG	11220
	TTTACACCAT	ATTTTTTACT	TACCGTTTTG	GCAGACGGCA	CTTGGTGGTA	CTTTAGAAGT	11280
	CAAAGGGCAC	TTAGTTCAAG	GTACGCAGAA	CATCTTCTTT	GCTCAACTTG	GTGATCCAGA	11340
20	TGTGACGAAG	TATTATTCAG	GTGTGTCACG	CTTTATGTCA	GGCCGTTTTA	TTACGATGAT	11400
	GTTCCGGCTTA	TGTGGTGCCG	CACCTGCAAT	TTATCACACA	GCTAAACCTG	AACATAAAAA	11460
	AGTTGTCCGC	GGTTTAATGT	TATCCGCTGC	ACTCACTTCA	TTTTTAACAG	GTATTACCGA	11520
25	ACCTTTAGAG	TTTAGTTTCT	TGTTTGTGCG	ACCTATTCTT	TATGTAATCC	ATGCCTTCTT	11580
	TGATGGATTA	GCATTTATGA	TGGCAGACAT	TTTCAACATT	ACAATTGGTC	AAACCTTCAG	11640
	TGGAGGCTTT	ATCGATTTCT	TACTCTTTGG	TGTGCTACAA	GGTAATAGTA	AAACAAACTA	11700
30	CCTATACGTC	ATACCTATTG	GAATTGTGTG	GTTCTGTTTG	TATTACATCG	TTTTCAGATT	11760
	CTTAATTACG	AAATTTAATT	TCAAAACACC	TGGTCGAGAA	GATAAAGCTG	CAGCACAACA	11820
	AGTTGAGGCT	ACTGAAAGAG	CACAAACTAT	TGTTGCTGGT	TTGGGAGGCA	AAGATAACAT	11880
35	TGAAATCGTT	GACTGTTGTG	CAACGAGACT	ACGCGTCACA	CTTCATCAAA	ATGACAAAGT	11940
	CGATAAAGTA	TTACTCGAAA	GTAAGGTGTC	CAAAGGTGTA	ATCCAGCAAG	GCACTGGTGT	12000
40	GCAAGTAATT	TATGGGCCTC	ACGTTACAGT	TATCAAAAAT	GAAATTGAAG	AATTGCTCGG	12060
	GGATTAAGAC	TAACCGAAAT	ATCAACAGAA	CTAATGGCAA	CGATGTACGA	AGTAAGAAGT	12120
	GACATCGTTG	CTTTTATTTT	TAATGTTACA	TTTGAAGCAT	TAAGTTCATC	ATGCACTGTA	12180
45	GTGAGCCCGC	AAATCGCCTC	TGCTAGACAA	TCATCTTAAT	GCTATGATTA	AAGCTTAAGT	12240
	GCCAGATTTG	AATTTAATTT	CAACAACGAC	TTTCACTACA	TTAAAAATAG	GGCCACTCGA	12300
	CACATATAGT	TGTATCAAAAT	AGCCCTTTAT	ACAATTTTTT	GGGTAAGGTT	TTACAATTTT	12360
50	TGGGATGGTA	TAGATTTTAT	AAAAAGTTAT	TTAAGTTCTT	CTGCTTCAGC	CATAATATCT	12420
	TTTAATGTTT	TAGCTGAATG	TGCGAACTTG	CTTTGTTCTT	CGTCGTTTAA	TGGGATTTCT	12480

55

	TCCTCATATT CGCCTTCTAA TAATGCTGAT ACAGTCAATA CGGCATCTTC ATTTCTGAAA	12600
	ATCGCTTCAG TAATTCCTAGC TAATCCCATT GCAACACCAT AATAAGTGGC ACCTTTAGCT	12660
5	TGAATAATGT CATATGCTGC ATCACGTGTT TGAACAAAAA TTTGTTCAAT TTGCGCTTTG	12720
	CCCTCAGGAC GTTGTTC AAG	12780
10	CATACTGGTA ATTCAGTGTC ACCATGTTCA CCAATAATTT GAGCATCGAC GCTACGTGGC	12840
	GCAACATCGn AcgyTcGCTT AACAATAATC TAAAGCGTGC AGAGTCTAAA ATTGTACCAG	12900
	AACCTATAAC ACGTCTCTTA GGTAAACCAG AGAATTTCCA TGTGCATAC GCTAAAATAT	12960
15	CAACAGGATT TGTAGCTACC AAGAAAATAC CATCAAATTT TGATGCCATT ACTTCACCAA	13020
	CAATTGATTT GAATATTTTC AAGTTTTTAG ATACTAAATC TAAACGTGTT TCTCCAGGTT	13080
	TTTGTGCAGC ACCAGCACAG ATGACAACCTA GATCCGCATC ATGACAATCA CTGTATTGCG	13140
20	CAGCTTTCAC ACGAACTGTT GTTGGAGAAT ATGGTGTGGC ATGTTTTTAAA TCCATAACAT	13200
	CTCCTCGAAC TTTTTCAGTG TCTAAATCAA TGATGACTAA TTCATCAACA ATGCTTTGGT	13260
	TCACTAATGA AAATGCGTAG CTTGAACCTA CTGCACCATT ACCTATTAAT ACAACTTTGT	13320
25	TCCCTTTAAA TTTGTTTATT AAAAAAATC CCTTATGATT AATTCATAA CATACTGTA	13380
	GCTTCAAATA TGTTAGTTTA ATGCTGCTTA TTGACCATAC AAAAGCAAAT AAACATCTCT	13440
	TTTATTTTCA ACGCATAACT TAAAAGGTCA TGTGTCATCC GCTTTTAAGT TTGTGATTTA	13500
30	TTTCACATAT AAAATGTAAC ATGCATTAAG TACTGGGTCA ATATTAAATT GTGATTTATT	13560
	TCACATTTTA TTTTAATTTT TACACCTTTT TAATTTGTAT mCGATTACAT CTTAGATGTC	13620
35	TTTAGTCTTC GTAATTCGCC AGTGATTATT TACACTTTCA CATTTTATT ATCATGTTTA	13680
	CTTTTTTCTA GGAAAAACAAC AATGTTTTTT GAATTAGTCA AATAAATGCG CTCAATCGTC	13740
	GGTGTGCAAA CAGACAATTG TACACAATGC TTATTGATAA GTATTTAAAA AATTAAAAAT	13800
40	GTCATACAAT TATCAAATTT GCCATTTTAT TTATATTTTC TCAAACCAAT TAATTGAATA	13860
	TCGAAATTTT TAGTAGAATA ATCAAAATAT ACAGATTAAA GGAGGAGTAT CATGCTTACA	13920
	GAACAAGAGA AAGACATTAT CAAACAAACG GTGCCTTTAC TTAAAGAGAA AGGGACAGAA	13980
45	ATTACGTCAA TCTTTTATCC AAAAATGTTT AAAGCGCATC CTGAACTTTT AAACATGTTT	14040
	AATCAAACGA ACCAAAAACG AGGCATGCAA TCTTCAGCAT TAGCACAAGC TGTAATGGCC	14100
	GCAGCGGTTA ATATCGATAA CTTAAGTGTT ATTAAACCAG TCATTATGCC AGTCGCATAT	14160
50	AAACACTGCG CACTACAAGT TTATGCTGAA CATTATCCAA TTGTGGGGAA AAATTTATTA	14220
	AAAGCCATTC AAGACGTGAC AGGATTAGAA GAAAATGACC CTGTCATTCA AGCTTGGGCA	14280
55		

## (2) INFORMATION FOR SEQ ID NO: 58:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 8779 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

GGTATTTTnG GAnGGGTACC TAAAGCAATT CCGGCAAAGG GTnAATCCAG GTACCGAAAT 60

GGACTTCCCG TTATCGATAA TACCGACATA TATTGTGACA AGTAGATTTT ATGGACATTT 120

AGGCTTACTT TTA CTGTGA TAATTGCATG TATGTTTACT GGTATTTAtC CaTCaATACA 180

TATCATTCAA TTATTGATAT ATGTACCGTT TTGTTTTTTC TTA ACTGCct CGGTGACGTT 240

ATTAACATCA ACAC TCGGTG TGTTAGTTAG AGATACACAA ATGTTAATGC AAGCAATATT 300

AAGAATATTA TTTTACTTTT CACCAATTTT GTGGCTACCA AAGAACCATG GTATCAGTGG 360

TTTAATTCAT GAAATGATGA AATATAATCC AGTTTACTTT ATTGCTGAAT CATACCGTGC 420

AGCAATTTTA TATCAGCAAT GGTATTTTCAT GGATCATTGG AAATTAATGT TATACAATTT 480

CGGTATTGTT GCCATTTTCT TTGCAATTGG TCGTACTTA CACATGAAAT ATAGAGATCA 540

ATTTCAGAC TTCTTGTAAT ATATTTATAT GACGAAACCC CGCTAACCAT TAATAAATGG 600

AAGTGGGGTT CATTTTTGTT TATAATTTAA GTAAATAACA TATTAAGTTG GTGTATTATG 660

AACGTTTTAA TAAAGAAATT TTATCATTG GTAGTTCGAA TACTTTCTAA AATGATTACG 720

CCTCAAGTGA TTGATAAACC GCATATCGTA TTTATGATGA CTTTTCCAGA AGATATTAAG 780

CCTATCATCA AAGCATTAAA TAATTCGTCG TATCAGAAAA CTGTTTTAAC AACACCAAAA 840

CAAGCGCCTT ATTTATCTGA ACTTAGCGAC GATGTTGATG TGATAGAAAT GACTAATCGA 900

ACATTGGTAA AACAAATTAA GGCTTTGAAA AGCGCGCAGA TGATTATTAT CGATAATTAT 960

TACCTATTGC TAGGTGGATA TAATAAGACT TCTAATCAAC ACATTGTTCA AACGTGGCAT 1020

GCAAGTGGTG CATTAAAAA CTTTGGCTTA ACAGATCATC AAGTCGATGT GTCTGACAAG 1080

GCAATGGTTC AGCAGTACCG TAAAGTTTAT CAAGCGACGG ATTTTTACTT AGTGGGTTGT 1140

GAACAAATGT CACAATGTTT TAAACAGTCT TTAGGTGCAA CAGAAGAGCA AATGCTGTAT 1200

TTTGGGCTTC CGAGAATTAA TAAATATTAC ACAGCTGATA GAGCAACGGT TAAGGCAGAG 1260

TTAAAGGATA AATATGGAAT TACAAATAAG TTGGTATTAT ATGTACCAAC ATATAGAGAA 1320

GATAAAGCAG ATAATAGGGC TATTGATAAA GCTTATTTTG AAAAATGTTT ACCAGGATAT 1380

	ATCGACACGT CTACATTAAT GCTAATGTCA GATATAATTA TTAGCGACTA TAGTTCGCTG	1500
	CCAATAGAAG CTAGCTTGTT AGATATTCCA ACTATATTTT ATGTGTATGA TGAAGGAACA	1560
5	TATGATCAGG TGAGAGGCCT GAATCAATTT TACAAAGCAA TACCGGATAG CTACAAAGTG	1620
	TATACTGAAG AAGATTTAAT AATGACGATA CAAGAAAAAG AACATCTATT AAGTCCGTTA	1680
	TTTAAAGATT GGCATAAGTA TAATACTGAT AAAAGTTTAC ATCAGCTCAC AGAATATATA	1740
10	GATAAGATGG TGACAAAATG AGGTTTACGA TAATCATACC TACATGTAAT AATGAGGCAA	1800
	CAATTCGACA ATTGTTAATA TCTATTGAGA GTAAAGAACA CTATAGAATC CTTTGTATTG	1860
	ATGGTGGTTC TACTGATCAA ACAATTCCTA TGATTGAACG GTTACAAAGA GAACTCAAGC	1920
15	ATATTTTCATT AATACAATTA CAAAATGCTT CGATAGCTAC GTGTATTAAT AAAGGTTTGA	1980
	TGGATATCAA AATGACAGAT CCACATGATA GTGACGCATT TATGGTCATA AAACCAACAT	2040
20	CAATCGTATT GCCAGGTAA TTAGATAGGT TAACTGCTGC TTTCAAAAAT AATGATAATA	2100
	TTGATATGGT AATAGGGCAG CGAGCTTACA ATTACCATGG TGAATGGAAA TTGAAAAGTG	2160
	CTGATGAGTT TATTAAAGAC AATCGAATCG TTACATTAAC GGAACAACCA GATTTGTTAT	2220
25	CAATGATGTC TTTTGACGGA AAGTTATTCA GTGCTAAATT TGCTGAATTA CAGTGTGaCG	2280
	AAACTTTAGC TAACaCATAC AATCACGCAA TACTTGTCaA GGCGATGCAA AAAGCTACGG	2340
	ATATACATTT AGTTTCACAG ATGATTGTGCG GAGATAACGA TATAGATACA CATGCTACAA	2400
30	GTAACGATGA AGATTTTAAT AGATATATCA CAGAAATTAT GAAAATAAGA CAACGAGTCA	2460
	TGGAAATGTT ACTATTACCT GAACAAAGGC TATTATATAG TGATATGGTT GATCGTATTT	2520
	TATTCAATAA TTCATTAAAA TATTATATGA ACGAACACCC AGCAGTAACG CACACGACAA	2580
35	TTCAACTCGT AAAAGACTAT ATTATGTCTA TGCAGCATTG TGATTATGTA TCGCAAAACA	2640
	TGTTTGACAT TATAAATACA GTTGAATTTA TTGGTGAGAA TTGGGATAGA GAAATATACG	2700
40	AATTGTGGCG ACAAACATTA ATTCAAGTGG GCATTAATAG GCCGACTTAT AAAAAATTCT	2760
	TGATACAACT TAAAGGGAGA AAGTTTGCAC ATCGAACAAA ATCAATGTTA AAACGATAAC	2820
	GTGTACATTG ATGACCATAA ACTGCAATCC TATGATGTGA CAATATGAGG AGGATAACTT	2880
45	AATGAAACGT GTAATAACAT ATGGCACATA TGACTTACTT CACTATGGTC ATATCGAATT	2940
	GCTTCGTCGT GCAAGAGAGA TGGGCGATTa TTTAATAGTA GCATTATCAA CAGATGAATT	3000
	TAATCAAATT AAACATAAAA AATCTTATTA TGATTATGAA CAACGAAAAA TGATGCTTGa	3060
50	ATCAATACGC TATGTCGATT TAGTCATTCC AGAAAAGGGC TGGGGACAAA AAGAAGACGA	3120
	TGTCGAAAAA TTTGATGTAG ATGTTTTTGT TATGGGACAT GACTGGGAAG GTGAATTCTGA	3180

55

	TAAAATCAAA CAAGAATTAT ATGGTAAAGA TGCTAAATAA ATTATATAGA ACTATCGATA	3300
	CTAAACGATA AATTAACCTA GGTATTATA AAATAAATAT AAAACGGACA AGTTTCGCAG	3360
5	CTTTATAATG TGCAACTTGT CCGTTTTTAG TATGTTTTAT TTTCTTTTTC TAAATAAACG	3420
	ATTGATTATC ATATGAACAA TAAGTGCTAA TCCAGCGACA AGGCATGTAC CACCAATGAT	3480
	AGTGAATAAT GGATGTTCTT CCCACATACT TTTAGCAACA GTATTTGCCT TTGAATAAT	3540
10	TGGCTGATGA ACTTCTACAG TTGGAGGTCC ATAATCTTTA TTAATAAATT CTCTTGGATA	3600
	GTCCGCGTGT ACTTTACCAT CTTGACTAC AAGTTTATAA TCTTTTTTAC TAAAATCACT	3660
	TGGTAAAACA TCGTAAAGAT CATTTTCAAC ATAATATTTT TTACCATTTA TCCTTTGCTC	3720
15	ACCTTTAGAC AATATTTTTA CATATTTATA CTGATCAAAT GAGCGTTCCA TTAATGCATT	3780
	CCCCATCATA TTACGTTGCT TCTCGCCACC AAGGTTTTTA TAGTCTCCTG CACCCATGAT	3840
	AACCTGATTA ATTCTAAATT TACCTCGTTT GGTAAGTAATC GTATGGTTGT AATTTGCTGT	3900
20	ATCACTTGAT CCAGTTTTTA AACCATCTGT ACCCGGCAAA CTCATTTTTG CACCTTCCAA	3960
	TGAAAAGTTG AATGTGTAAT ACGTAACTGC ATGCGTTGTT GGTGCTAACT GCTTTGTAAA	4020
25	GTCTAATATT TTAGGTGTCT CTTTAATCAC GTGTAAATCT AAAATGGCAT AGTCTCTAGC	4080
	AGTCGTTACA GTACGTTCTT GGTCTTTATA CTTTGTTGGT GCAAATGTAC GTAATCTTGA	4140
	ATTTTCAGCA CCCGTTGGAT TGACGAAATG TGTATTTTTT ATTCCGATAG CTTTAGCTTT	4200
30	GTTATTCATT AAATCAACGA AATCGCTGGT GTTTTTTGAA ACCTTCTTAG CTAAAATTA	4260
	TGCCGCGGCA TTAGTAGAAT TAGATACTGT AATTGTGAAT AGGTCTGCGA TTGTCCATAC	4320
	TTGTCCAGGA TATAGTTTCG TATTACTCAA CTCAGGTAGT GTAGACATAA TATATCTTT	4380
35	GTTCGTCATT GTGACTGTGT CATCAAGTGA AAGCTGCCCC TTATTTACAG CTTCCAATGT	4440
	TAAGTACATT GTCATTAATT TAGTCATAGA CGCTGGATTC CACTTAGTAT CGATATTGTA	4500
	TTGATACAGT AATTGTCCAG TTTGACTTAC ATTAACAGCA CTCGTCGGTT CGTATGCAGC	4560
40	CGACAAACCT GCATAACCAT ATTGATTTGC TGCTTGTAACA GGGGTTACGT CACTGTTAGT	4620
	AGCTTGTCGA TATGGTGTCA TAATACTTAA TGTTAAACAT AAAATGATGA TAATAGATAT	4680
	TAAATTTTTT ATAAAGCGTT AATCTTCCCT TTTCCAATTC TTAAATATTC CCTAAAAGCA	4740
45	ATGGTTATTC CTACTTACGG AAATCATTGC TAATTCATT CACCTTAATT AAATTGTTGA	4800
	AAATAAGTT TTCTGCAGTT AATTGAAAA ATAATGCAAA TATATTACGT GTGTAGCTAA	4860
50	AGGTGTTATA ATGTTTGTAC GAAGAGCAAA CTTACTCAA AGCGATTAAAT TTTTCATGTTT	4920
	TAATATAAAG ACTTTGAGAA GTTATTACAA AAAATGCAAT AGAAATATTC TATCATATAA	4980

55

	AAGTATATGA TAGAAATGCA TGTATCTATC TAAATGAATT AACTATAAAT TTCAAACAGA	5100
	AGAGGTAAAA CTATGAAACG AGAAAATCCA TTGTTTTTCT TATTTAAAAA ACTATCATGG	5160
5	CCAGTGGGTC TTATCGTTGC AGCTATCACT AITTCATCAC TAGGGAGCTT AAGTGGACTA	5220
	TTAGTGCCAC TGTTTACTGG ACGAATTGTA GATAAATTTT CCgTGAGCCA TATCAATTGG	5280
	AATCtAATCG CATTATTTGG TGCTATCTTT GTCATCAATG CTTTATTAAG CGGATTAGGT	5340
10	TTATATTTAT TAAGTAAAT TGGTGAAAAG ATTATTTATG CGATACGCTC AGTTTTATGG	5400
	GAGCATATCA TACAATTAA AATGCCATTC TTTGACAAAA ATGAAAGTGG TCAATTAATG	5460
15	AGTCGATTAA CTGACGATAC GAAAGTGATA AATGAATTTA TTTCACAAAA GCTACCTmAC	5520
	TTATTACCAT CAATCGTTAC ATTAGTTGGG TCACTAATCA TGTTATTTAT TTTAGATTGG	5580
	AAAATGACAT TATTAACATT TATAACGATA CCGATATTCG TTTTaATTAT GATTCCTCTA	5640
20	GGTCGTATTA TGCAAAAGAT ATCGACAAGT ACACAATCTG AAATTGCAAA CTTCAGTGGT	5700
	TTGTTAGGGC GTGTCCTAAC TGAAATGCGT CTTGTTAAAA TATCAAATAC AGAGCGTCTT	5760
	GAATTAGATA ATGCACATAA AAATTTGAAT GAAATATATA AATTAGGTTT AAAACAGGCT	5820
25	AAAATTGCGG CAGTTGTACA ACCAATTTC A GGTATAGTTA TGTTGCTAAC AATTGCAATT	5880
	ATTTTAGGTT TTGGTGCATT AGAAATTGCG ACTGGTGCAA TCACTGCAGG TACATTAATT	5940
	GCAATGATAT TTTATGTTAT TCAGTTATCT ATGCCTTTAA TCAATCTTTC CACGTTAGTT	6000
30	ACAGATTATA AAAAGGCAGT CCGTGCAAGT AGTAGAATAT ACGAAATCAT GCAAGAACCT	6060
	ATTGAACCGA CAGAAGCTCT TGAAGATTCT GAAAATGTAT TAATTGATGA CGGTGTATTG	6120
	TCATTTGAAC ATGTAGACTT TAAATATGAT GTGAAGAAAA TATTAGATGA TGTGTCGTTT	6180
35	CAAATCCAC AAGGTCAAGT GAGTGCTTTT GTAGGCCCTT CTGGGTCTGG TAAAAGTACG	6240
	ATATTTAATC TGATAGAACG TATGTATGAA ATTGAGTCAG GTGATATTAA ATATGGCCTT	6300
	GAAAGTGTCT ATGATATCCC GTTATCTAAG TGCGCAGCA AAATTGGATA TGTTATGCAA	6360
40	TCAAATTCGA TGATGAGTGG TACAATTAGA GACAATATTT TATACGGAAT TAATCGTCAT	6420
	GTTTCAGATG AAGAACTTAT TAATTATGCT AAATTAGCGA ACTGTCATGA TTTTATCATG	6480
45	CAATTTGATG AAGGATATGA CACGCTTGTA GGTGAACGAG GATTGAAACT GTCTGGCGGA	6540
	CAACGTCAAC GTATTGATAT TGCTAGAAGT TTTGTTAAAA ATCCTGATAT TTTGTTACTT	6600
	GATGAAGCAA CAGCTAATCT CGATAGTGAA AGTGAATTGA AAATTCAAGA AGCTTTAGAA	6660
50	ACATTGATGG AAGGTAGAAC AACGATTGTC ATTGCGCATC GTTTGTCTAC AATTAAAAAA	6720
	GCCGGTCAAA TTATATTCTT AGACAAAGGA CAGGTAACAG GTAAAGGTAC GCATTCAGAA	6780

55



	TTTTATATAT ATAAGTAAGC TTGGAGCAAA TACACATATA CCATCGAGGA AATTAAAGTG	6900
	TGGCACATTG ATGGATATAG ATGTTAATAA ATTGCTTCAA GCTTTTGTCT ATTTTAAATC	6960
5	ATTTGAGAAG TTACGACATA ATAATTCTTA AATTAATGAA ATCGATATTT TAAGAAAAAA	7020
	ATGCTCATGG TATAATACAA GTTATAAGCA AACATACATA TATTAAATAC TGTAGCCACG	7080
	AGTCATAATT CTTCATATTT TACATAGCAA TTAACTGAT TTTAGAGTCC ACGGTACAGA	7140
10	AGTTTGATAT TTCAATGTTT CTAAATTTTT AAAAAATTAA ATCATAGGTG GGTGCCAAAT	7200
	GTTTTTATTA ATCAACATTA TTGGTCTAAT TGTATTTCTT GGTATTGCGG TATTATTTTC	7260
	AAGAGATCGC AAAAATATCC AATGGCAATC AATTGGGATC TTAGTTGTTT TAAACCTGTT	7320
15	TTTAGCATGG TTCTTTATTT ATTTTGATTT GGGTCAAAAA GCAGTAAGAG GAGCAGCCAA	7380
	TGGTATCGCT TGGGTAGTTC AGTCAGCGCA TGCTGGTACA GGTTTTGCAT TTGCAAGTTT	7440
	GACAAATGTT AAAATGATGG ATATGGCTGT TGCAGCCTTA TTCCAATAT TATTAATAGT	7500
20	GCCATTATTT GATATCTTAA TGTACTTTAA TATTTTACCG AAAATTATTG GAGGTATTGG	7560
	TTGGTTACTA GCTAAAGTAA CAAGACAACC TAAATTCGAG TCATTCTTTG GGATAGAAAT	7620
25	GATGTTCTTA GGAAATACTG AAGCATTAGC CGTATCAAGT GAGCAACTAA AACGTATGAA	7680
	TGAAATGCGT GTATTAACAA TCGCAATGAT GTCAATGAGC TCTGTATCGG GAGCTATTGT	7740
	AGGTGCGTAT GTACAAATGG TACCAGGAGA ACTGGTACTA ACGGCAATTC CACTAAATAT	7800
30	CGTTAACGCG ATTATTGTGT CATGCTTGTT GAATCCAGTA AGTGTGAAG AGAAAGAAGA	7860
	TATTATTAC AGTCTTAAAA ACAATGAAGT TGAACGTCAA CCATTCTTCT CATTCCTTGG	7920
	AGATTCTGTA TTAGCAGCAG GTAAATTAGT ATTAATCATC ATCGCATTTG TTATTAGTTT	7980
35	TGTAGCGTTA GCTGATCTAT TTGATCGTTT TATCAATTTG ATTACAGGAT TGATAGCAGG	8040
	ATGGATAGGC ATAAAAGGTA GTTTCGGTTT AAACCAAATT TTAGGTGTGT TTATGTATCC	8100
	ATTTCGCTA TTA CTGGTT TACCTTATGA TGAAGCGTGG TTGGTAGCAC AACAAATGGC	8160
40	TAAGAAAATT GTTACAAATG AATTTGTTGT TATGGGTGAA ATTTCTAAAG ATATTGCATC	8220
	TTATACACCA CACCATCGTG CGGTTATTAC AACATTCTTA ATTCATTTG CAAACTTCTC	8280
45	AACGATTGGT ATGATTATCG GTACATTGAA AGGCATTGTT GATAAAAAGA CATCAGACTT	8340
	TGTATCTAAA TATGTACCTA TGATGCTATT ATCAGGTATC CTAGTTTCAT TATTAACAGC	8400
	AGCTTTCGTT GGTTTATTTG CATGGTAATA TGTCGAAGAG TGA CTATGAT AATACATTTT	8460
50	AACTAATAAA TATGTCCAGG CATGTCGTCT ATTGATATAG GTGAGATGCT TGGACTTTTT	8520
	TATTATTGAT ATAAAGGTAT nTAAATATTT TTAAAGTTAC CGAAATTGAA GCATTATAAA	8580

55

GACAGTAAGG ACTAGGTACA GTCATAGTAC TTCGAGCAAA ATTTGTTTTG TTATTATAAA 8700  
 CAACACAAAAG GAGATAACTT CTCTAaTGAA GAAGTTAAAA ACATTATAGC AGACAATGAA 8750  
 5 ATGAAAGTAA ATTAAAAAT 8779

(2) INFORMATION FOR SEQ ID NO: 59:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31096 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

GTTGCAGTAG TCAAAGAATT AAACAAGGTG AAGGcGTGTA GCTTGCACAC CCGAAAATGT 60  
 20 GCGTAAGTTA aCGGATGCAG GACATAAAGT AATTGTTGAA AAAAATGCTG GCATTGGTTC 120  
 AGGATTTTCT AACGATATGT ATGAAAAAGA AGGCGCTAAG ATCGTAACTC ACGAACAAGC 180  
 ATGGGAAGCT GATCTTGTTA TCAAAGTAAA AGAACCTCAT GAAAGCGAAT ATCAATATTT 240  
 25 CAAAAAGAAT CAAATTATCT GGGGATTTT ACATCTAGCA TCTTCAAAAG AAATAGTAGA 300  
 AAAAATGCAA GAAGTTGGTG TAACTGCGAT TAGTGGTGAA ACCATTATAA AAAATGGAAA 360  
 AGCAGAATTA TTAGCGCCAA TGAGTGCTAT AGCAGGTCAA CGCTCAGCAA TTATGGGAGC 420  
 30 TTACTACTCT GAAGCACAAAC ATGGTGGTCA AGGTACTTTA GTGACTGGTG TACATGAAAA 480  
 TGTGGATATA CCTGGTAGTA CATATGTGAT TTTCGGTGGT GGAGTAGCAG CAACAAATGC 540  
 AGCAAATGTT GCCTTGGGAC TAAATGCTAA AGTAATCATT ATCGAGTTAA ACGATGACCG 600  
 35 CATTAATAT CTTGAAGATA TGTATGCAGA AAAAGATGTC ACAGTAGTCA AATCAACACC 660  
 AGAA<sup>-</sup>AATTTA GCAGAACAAA TTAAGAAAGC AGATGTATTT ATTTCTACAA TTTTAATTTT 720  
 40 AGGTGCGAAA CCGCCAAAAT TGGTTACTCG TGAGATGGTT AAATCAATGA AAAAAGGTTC 780  
 AGTATTAATC GATATAGCTA TTGACCAAGG TGGAACATT GAAACAATTA GACCAACTAC 840  
 AATTTCTGAT CCAGTGTATG AAGAAGAAGG TGTGATTCAT TATGGTGTAC CAAATCAACC 900  
 45 AGGAGCAGTC CCAAGAACTT CAACAATGGC ATTAGCACAA GGAAATATTG ATTATATATT 960  
 AGAAATTTGT GACAAAGGCT TAGAACAAGC AATTAAAGAT AATGAAGCCT TAAGTACTGG 1020  
 TGTAACATT TACCAAGGAC AAGTGACAAA TCAAGGATTA GCTTCATCAC ATGACCTAGA 1080  
 50 TTATAAGAA ATATTAAATG TTATCGAATA GATAGTAATT TAAATGAAAT TGAGTGAAAT 1140  
 GAATATTTTA AATATAGCAT TATAGTTTGG ACTAAAAATT TACAAAACGG AAGGATGTAA 1200

	TCGAAGAAGC TAAAGCAAGC ATTAAACCAT TTATTCGTCG AACACCTCTA ATTAAATCAA	1320
	TGTATTTAAG CCAAAGTATA ACTAAAGGGA ATGTATTTCT AAAATTAGAA AATATGCAAT	1380
5	TCACAGGATC TTTTAAATTT AGAGGCGCTA gCAATnAAAA TTAATCACTT AACAGATGAA	1440
	CAAAAAGAAA AAGGCATTAT CGCAGCATCT GCTGGGgAAC CATGCACAAG GTGTTGCTTT	1500
	AACAGCTAAA TTATTAGGCA TTGATGCAAC GATTGTAATG CCTGAAACAG CACCACAAGC	1560
10	GAAACAACAA GCAACAAAAG GCTATGGGGC AAAGGTTATT TTAAAAGGTA AAAACTTTAA	1620
	CGAAACTAGA CTTTATATGG AAGAATTAGC GAAAGAAAAT GGCATGACAA TCGTTCATCC	1680
	ATATGACGAT AAGTTTGTAA TGGCAGGCCA AGGAACAATT GGTTTAGAAA TTTTAGATGA	1740
15	TATTTGGAAT GTGAATACAG TCATCGTACC AGTTGGCGGT GGAGGATTAA TTGCAGGTAT	1800
	TGCCACCGCA TTAAAATCAT TTAACCTTC AATTCATATT ATCGGTGTTT AATCTGAGAA	1860
20	TGTTTCATGGT ATGGCTGAGT CTTTCTATAA GAGAGATTTA ACTGAACATC GAGTGGATAG	1920
	CACAAATAGCA GATGGTTGTG ATGTAAAAGT TCCTGGTGAA CAAACATATG AAGTAGTTAA	1980
	ACATTTAGTA GATGAATTTA TTCTTGTTAC TGAAGAAGAA ATTGAACATG CTATGAAAGA	2040
25	TTTAATGCAG CGTGCCAAAA TTATTACTGA AGGTGCAGGC GCATTACCAA CAGCTGCAAT	2100
	TTTAAGTGGA AAAATAAACA ATAAATGGCT TGAAGATAAA AATGTTGTTG CATTAGTTTC	2160
	AGGCGGGAAT GTTGACTTAA CTAGAGTTTC AGGTGTCATT GAACATGGAC TGAATATTGC	2220
30	AGATACAAGC AAGGGTGTGG TAGGTAAAAA CATTTAATCT TAAAAATGAG GTGTAATTAT	2280
	GTCAAATGGT AAAGAATTAC AAAAAATAT AGGTTTCTTC TCAGCGTTTG CTATTGTTAT	2340
	GGGGACAGTT ATTGGTTCAG GAGTATTCTT TAAAATATCA AACGTAACAG AAGTAACAGG	2400
35	AACAGCAGGA ATGGCCTTGT TTGTATGGTT CCTAGGCGGC ATCATTACCA TTTGTGCGGG	2460
	GTTAACAGCA GCAGAACTTG CTGCTGCAAT CCCTGAAACA GGTGGCTTAA CGAAGTATAT	2520
	AGAATATACA TACGGTGATT TCTGGGGCTT CCTATCAGGT TGGGCGCAAT CATTATTTTA	2580
40	TTTTCAGCT AACGTAGCAG CATTGTCTAT CGTATTTGCG ACACAGCTAA TTAATTTATT	2640
	CCATTTATCT ATAGGTTGCT TAATACCAAT AGCAATCGCA TCTGCGTTAT CTATTGTGTT	2700
45	GATAAATTTT CTAGGTTCAA AAGCAGGCGG AATTTTACAA TCAGTTACTT TAGTAATTAA	2760
	ACTGATTCCA ATCATCGTTA TTGTAATTTT TGGTATTTT CAATCTGGAG ATATCACTTT	2820
	TTCATTAATT CCAACTACAG GTAATTCaGG AAATGGCTTC TTTACAGCAA TTGGTAGTGG	2880
50	TTTATTAGCA ACTATGTTTG CATATGATGG TTGGATTCAT GTAGGAAATG TTGCGGGGA	2940
	ACTTAAAAAT CCTAAACGCG ATTTACCTTT AGCGATTTCA GTTGGTATCG GTTGTATTAT	3000

55

	TGGTAATTTA AATGCAGCTT CAGATACATC AAAAAATATTA TTTGGTGAAA ATGGCGGTAA	3120
	GATTATTACA ATCGGTATAT TAATTTCTGT TTATGGTACG ATCAATGGCT ATACTATGAC	3180
5	TGGTATGCGC GTACCATATG CAATGGCTGA AAGAAAATTA TTGCCATTTA GCCATTTATT	3240
	CGCAAAATTA ACAAATCTG GCGCACCATG GTTTGGCGCA ATTATACAAC TTATAATCGC	3300
	TATCATCATG ATGTCAATGG GAGCATTTGA TACAATTACA AATATGTTAA TCTTTGTTAT	3360
10	TTGGTTGTTT TATTGTATGT CATTGTGTGC GGTAATAATT TTAAGAAAAC GTGAACCAAA	3420
	TATGGAACGA CCATATAAAG TACCGTTATA TCCGATCATA CCTTTAATTG CTATTTTGGC	3480
	AGGATCATTT GTATTAAATTA ATACACTGTT TACACAATTT ATATTAGCAA TCATTGGAAT	3540
15	TCTAATAACA GCACCTGGTA TACCAGTTTA TTACTATAAA AAGAAACAAA AAGCAGCATA	3600
	AGGTAAGATA ACTAGCATTG AGAATAAATG GATGGACTAC TAATAAATTT AAAGTTTTAC	3660
20	ACATTAAAT CAAAAACCAT TCAATTATTC TATGGAACAG ACAAATTTCT GTTATGGAAT	3720
	TTGTCTGTTT TTCAAAGTA TAGGGAGGCA AATAGAGATG GAAAAGCCGT CAAGAGAGGC	3780
	ATTTGAAGGC AATAATAAGT TGTTAATAGG AATTGTTCTA AGTGTAATAA CGTTTTGGCT	3840
25	ATTTGCACAA TCATTGGTTA ATGTTGTACC AATACTTGAA GATAGTTTCA ATACAGATAT	3900
	TGGAACGGTT AATATCGCCG TTAGTATAAC TGCTTTATTT TCAGGAATGT TTGTAGTAGG	3960
	AGCAGGTGGT CTTGCTGATA AATATGGCAG AATTAACTC ACGAACATTG GTATTATCTT	4020
30	AAATATATTA GGTTCAATTAT TAATCATTAT TTCAAATATT CCTTTATTAC TTATTATAGG	4080
	AAGATTAATT CAAGGACTTT CAGCAGCATG TATTATGCCT GCAACTTTGT CTATTATTAA	4140
	GTCATATTAC ATTTGGGAAAG ATAGACAACG CGCTTTAAGT TATTGGTCAA TTGGCTCATG	4200
35	GGGCGGCTCT GGTGTTTGTT CATTTTTTGG AGGTGCAGTT GCAACGCTTT TAGGTTGGCG	4260
	TTGGATTTTC ATCCTATCAA TTATAATTTT ATTAATTGCA CTGTTTCTTA TTAAAGGCAC	4320
	ACCTGAAACT AAATCTAAAT CGATTTCTCT AAATAAATTT GACATTAAAG GTCTGGTTCT	4380
40	TTTAGTCATT ATGCTCCTCA GTTTAAATAT TTTAATTACT AAAGGATCAG AATTAGGTGT	4440
	AACCTCACTT CTTTTTATTA CTTTATTAGC TATTGCAATT GGATCTTTTA GTTTATTTAT	4500
45	AGTTCTTGAA AAGCGTGCTA CAAATCCTTT AATCGATTTT AAATTATTTA AAAATAAAGC	4560
	TTACACAGGT GCAACAGCTT CAACTTTTTT GTTAAATGGT GTTGCAGGAA CATTAATAGT	4620
	AGCCAACACA TTTGTTCAAA GAGGTTTAGG ATATTCTTCA TTGCAAGCAG GAAGTTTATC	4680
50	AATCACTTAT TTAGTAATGG TACTAATTAT GATTCGTGTT GGTGAAAAGT TACTTCAAAC	4740
	ACTCGGATGC AAGAAACCAA TGTTAATTGG AACAGGAGTT CTTATTGTCTG GAGAATGTCT	4800

55

	ATTCTTTGGT	TTAGGACTAG	GGATATATGC	TACACCATCA	ACAGATACAG	CAATTGCAAA	4920
	TGCACCGTTA	GAAAAAGTAG	GCGTTGCTGC	AGGTATCTAT	AAAATGGCTT	CTGCATTAGG	4980
5	TGGAGCATTT	GGCGTCGCAT	TGAGTGGTGC	AGTATATGCA	ATCGTATCAA	ATATGaCAAA	5040
	CATTTATACA	GGTGcAATGa	TTGnCAATTAT	GGTTaAATGC	AGGTATGGGa	ATATTATCaT	5100
	TCGTTATCAT	TTTGtTACTT	GTGcCTAAAC	mAAACGACAC	TCAATTATGA	TAATTGAGAA	5160
10	TTAAATTGAA	ATCATACAAG	TCGCTACAAT	ATTAAACAAA	AATATAAACC	GATTCTTATG	5220
	TGTCATTATT	TTAAATGAAC	ATAGGGATTG	GTTTTTTATT	ACTCTTTTAC	GCTACTTTAT	5280
15	TTATAATTAT	TATAAAATTGT	CACAAATTCA	ATTTACCTTA	CAATATATTT	TGTGTTATTA	5340
	TATTCTGGAG	CATAAATAAA	TTGTTCAACA	CATAGTTGTA	ATGTGTTTCA	ATACTTTTTG	5400
	GATAGATTGC	GAAATTGTAT	TGAATCGTCA	TCGTTTTAAA	TTTTTAAATG	AGAATGGAAT	5460
20	GAGCATTACA	ATACACAAGC	AATCAAAAAGT	AAATACATTC	ACAACACAAC	AGAGACATAA	5520
	CAACAAGATA	AGGAGTGAAC	AATAGCTGTG	AATTATCGTG	ATAAAATTCA	AAAGTTTAGT	5580
	ATTCGTAAAT	ATACAGTTGG	TACATTTTCA	ACTGTCATTG	CGACATTGGT	ATTTTTAGGA	5640
25	TTCAATACAT	CACAAGCACA	TGCTGCTGAA	ACAAATCAAC	CAGCAAGCGT	GGTTAAACAG	5700
	AAACAACAAA	GTAATAATGA	ACAGACTGAG	AATCGAGAAT	CTCAAGTACA	AAATTCTCAA	5760
	AATTCACAAA	ATGGTCAATC	ATTATCTGCT	ACTCATGAAA	ATGAGCAACC	AAATATTAGT	5820
30	CAAGCTAATT	TAGTAGATCA	AAAAGTAGCG	CAATCATCTA	CTACTAATGA	TGAACAACCA	5880
	GCATCTCAAA	ATGTAAATAC	AAAGAAAGAT	TCGGCAACGG	CTGCGACAAC	ACAACCAGAT	5940
35	AAAGAACAAA	GTAAGCATAA	ACAAAACGAA	AGTCAATCTG	CTAATAAAAA	TGGAACGAC	6000
	AATAGAGCGG	CTCATGTAGA	AAATCATGAA	GCAAATGTAG	TAACAGCTTC	AGATTCATCT	6060
	GATAATGGTA	ACGTACAACA	TGACCGAAAT	GAATTACAAG	CGTTTTTTGA	TGCAAATTAT	6120
40	CATGATTATC	GCTTTATTGA	CCGTGAAAAT	GCAGATTCTG	GCACATTTAA	CTATGTAAAA	6180
	GGCATTTTTG	ATAAGATTAA	TACGTTATTA	GGCAGTAATG	ATCCAATAAA	CAATAAAGAC	6240
	TTGCAACTTG	CATACAAAGA	ATTGGAACAA	GCTGTTGCTT	TAATTGCTAC	AATGCCTCAA	6300
45	CGTCAACAGA	CTAGCCGACG	TTCAAATAGA	ATTCAAACGC	GTTGCGTTGA	GTCAAGAGCT	6360
	GCAGAGCCTA	GATCAGTATC	AGACTATCAA	AATGCAAATT	CATCATATTA	TGTTGAAAA	6420
	GCTAATGATG	GTTGCGGCTA	TCCTGTTGGT	ACATATATCa	ATGCTTCTAG	TAAAGGGGCG	6480
50	CCATATAAIT	TACCAACTAC	ACCATGGAAT	ACATTGAAGG	CCTCTGACTC	AAAGGAAATT	6540
	GCTCTTATGA	CAGCGAAACA	AACTGGAGAC	GGGTACCAAT	GGGTTATTAA	GTTTAATAAA	6600
55							

## EP 0 786 519 A2

GTAGGAAGAA CTGACTTTGT AACAGTTAAT TCAGATGGAA CAAATGTACA ATGGAGTCAT 6720  
 GGAGCAGGAG CAGGTGCAAA TAAACCACTT CAACAAATGT GGGAATATGG AGTAAATGAT 6780  
 5 CCTCATCGTT CACATGACTT TAAAATAAGA AATAGAAGTG GCCAAGTAAT ATATGACTGG 6840  
 CCAACTGTCC ATATTTATTC TTTAGAAGAT TTATCTAGAG CGAGTGATTA TTTTAGTGAA 6900  
 GCTGGAGCGA CACCTGCTAC TAAAGCTTTT GGTAGACAAA ATTTTGAATA TATTAATGGT 6960  
 10 CAAAAACCTG CTGAATCACC GGGTGTTCCT AAAGTTTATA CTTCATCGG TCAAGGTGAT 7020  
 GCAAGTTATA CAATTTTATT TAAACACAA GGTCCAACCTG TTAATAAATT GTACTATGCA 7080  
 GCAGGTGGGC GTGCTTTAGA GTACAATCAA TTATTTATGT ACAGTCAACT ATACGTCGAA 7140  
 15 TCAACGCAAG ACCATCAACA ACGTCTTAAT GGTTTAAGAC AAGTGGTTAA TCGTACATAT 7200  
 CGCATAGGTA CAACTAAACG TGTAGAAGTG AGTCAAGGAA ATGTACAAAC GAAAAAGGTA 7260  
 TTAGAAAGTA CAAACCTAAA TATAGATGAT TTTGTTGATG ATCCTTTAAG TTATGTTAAG 7320  
 20 ACGCCGAGTA ATAAAGTGTT AGGATTTTAT TCGAATAATG CAAATACTAA TGCTTTTAGA 7380  
 CCGGGTGGAG CCCAACAATT AAATGAATAT CAATTAAGTC AATTATTTAC TGATCAAAAA 7440  
 25 TTACAAGAAG CAGCAAGAAC TAGAAACCCA ATAAGATTAA TGATTGGTTT CGACTATCCT 7500  
 GATGCTTATG GTAATAGTGA AcTTTAGTTC CTGTAACTT AACGGTATTA CCTGAAATCC 7560  
 AACATAATAT TaAATTCTTT AAAAATGACG ATACTCAAAA TATTGCTGAA AAACCATTTT 7620  
 30 CAAAAACAAGC TGGGCATCCA GTTTCTATG TATATGCAGG TAACCAAGGG AATGCTTCCG 7680  
 TGAATTTIAGG TGGTAGCGTA ACATCTATTC AACCATTACG TATTAATTTA ACAAGTAATG 7740  
 AGAATTTTAC AGATAAAGAT TGGCAAATTA CAGGTATTCC GCGTACATTA CACATTGAAA 7800  
 35 ACTCGACAAA TAGACCTAAT AATGCCAGAG AACGCAATAT TGAACCTGTT GGTAACTTAT 7860  
 TACCGGGGA TTACTTTGGA ACGATACGTT TTGGACGTAA AGAACAATTA TTCGAAATTC 7920  
 GTGTTAAACC ACATACACCA ACAATTACAA CGACAGCTGA GCAATTAAGA GGTACAGCAT 7980  
 40 TACAAAAAGT GCCTGTTAAT ATTTGCGGAA TACCGTTGGA TCCATCGGCA TTGGTTTATT 8040  
 TAGTTGCACC AACAAATCAA ACTACGAATG GTGGTAGTGA GGCAGATCAA ATACCATCTG 8100  
 GTTATACGAT ACTTGCGACT GGTACACCTG ATGGGGTGCA TAATACAATT ACTATACGAC 8160  
 45 CGCAAGATTA TGTGTATTC ATACCACCTG TAGGTAAACA AATTAGAGCA GTAGTTTATT 8220  
 ATAATAAAGT AGTTGCATCT AATATGAGTA ATGCTGTTAC TATTTTGCCA GATGACATTC 8280  
 50 CACCAACAAT CAATAATCCT GTTGAATAA ATGCCAAATA CTATCGAGGC GACGAACCAA 8340  
 CTTTACAATG GGTGTCTCTG ATAGACATTC TGGTATAAAA AATACAACCTA TTACGACATT 8400

55

	TACAGGTAGA GTGAGTATGA ATCAGGCATT TAACAGTGAT ATTACATTTA AAGTGTCAGC	8520
	GACAGaCAAT GTCAATAATA CGACAAATGA TAGTCAATCT AAACATGTTT CAATTCATGT	8580
5	AGGTAAAATT AGTGAAGATG CTCATCCGAT TGTATTAGGA AATACTGASA AAGTTGTAGT	8640
	AGTCAATCCG ACTGCTGTAT CTAATGATGA AAAGCAAAGC ATAATTACTG CCTTTATGAA	8700
	TAAAAACCAA AATATAAGAG GATATTTAGC ATCAACTGAT CCAGTAACTG TCGATAATAA	8760
10	TGGTAATGTC ACATTACATT ACCGTGATGG CTCATCGACA ACGCTTGATG CTACAAATGT	8820
	GATGACATAC GAACCAAGTTG TGAAACCTGA ATACCAAAGT GTCAATGCTG CTAAAACAGC	8880
15	AACGGTAACG ATTGCTAAAG GACAATCATT TAGTATTGGT GATATTAAAC AATATTTTAC	8940
	TTTAAGTAAT GGACAACCTA TTCCAAGTGG CACATTTACA AATATTACAT CTGATAGAAC	9000
	TATTCCAACG GCACAAGAAG TTAGTCAAAT GAACGCAGGC ACGCAGTTAT ACCATATAAC	9060
20	TGCTACAAAT GCGTATCATA AAGATAGTGA AGACTTCTAT ATTAGTTTGA AAATCATCGA	9120
	TGTGAAACAA CCAGAAGGCG ATCAACGTGT ATATCGTACA TCAACATATG ATTTAACTAC	9180
	TGATGAAATC TCAAAAGTAA AACAAGCATT TATTAATGCA AATAGAGATG TAATTACGCT	9240
25	TGCCGAAGGT GATATTTTCA TTACAAATAC ACCTAATGGT GCTAATGTAA GTACTATTAC	9300
	AGTAAATATT AATAAAGGTC GATTAACGAA ATCATTCGCG TCAAACCTAG CTAATATGAA	9360
	TTTCTTGCGT TGGGTTAATT TCCCACAAGA TTATACAGTG ACATGGACGA ATGCAAAAAT	9420
30	TGCAACAGA CCAACAGATG GTGGTTTATC ATGGTCTGAT GACCATAAAT CTTTAATTTA	9480
	TCGTTATGAT GCTACATTAG GTACTCAAAT TACGACGAAT GATATTTTAA CAATGTTAAA	9540
	AGCAACAACT ACAGTGCCTG GATTGCGAAA TAACATTACT GGTAATGAAA AATCACAAGC	9600
35	AGAAGCTGGC GGAAGACCTA ACTTTAGAAC GACTGGTTAT TCACAATCAA ATGCGACAAC	9660
	TGATGGTCAA CGTCAATTTA CGTTGAATGG TCAAGTGATT CAAGTGTTAG ACATCATCAA	9720
	CCCTTCAAAC GGTATGGTG GGCAACCTGT TACAAATTCA AATACTCGTG CAAACCATAG	9780
40	TAACTCAACT GTTGTTAACG TAAACGAACC GGCAGCTAAT GGTGcTGGCG CATTTACAAT	9840
	TGACCACGTT GTAAAAAGTA ATTCTACACA TAATGCAAGT GATGCAGTTT ATAAAGCACA	9900
45	GTTATACTTA ACGCCATATG GTCCAAAACA ATATGTTGAA CATTTAAATC AAAATACAGG	9960
	AAATACTACT GACGCTATTA ACATTTATTT TGTACCAAGT GACTTAGTGA ATCCAACAAT	10020
	TTCAGTAGGT AATTACACTA ATCATCAAGT GTTCTCAGGT GAAACATTTA CAAATACTAT	10080
50	TACAGCGAAT GATAACTTTG GTGTGCAATC TGTAAGTGTA CCAAATACAT CACAAATTAC	10140
	AGGTACTGTT GATAATAACC ATCAACATGT TTCTGCAACG GCACCAAATG TGACATCAGC	10200

55

EP 0 786 519 A2

	GTTCAATGTA ACAGTGAAAC CTTTGCGTGA TAAATATCGA GTTGGTACTT CATCAACGGC	10320
	TGCTAATCCT GTGAGAATTG CCAATATTTT GAATAATGCG ACAGTATCAC AAGCTGATCA	10380
5	AACGACAATT ATTAATTCGT TAACGTTTAC TGAAACAGTA CCAAATAGAA GTTATGCAAG	10440
	AGCAAGTGCG AATGAAATCA CTAGTAAAC AGTTAGTAAT GTCAGTCGTA CTGGAAATAA	10500
	TGCCAATGTg cACAGTAACT GTTACTTATC AAGATGGAAC AACATCAACA GTGACTGTAC	10560
10	CTGTAAAGCA TGTCATTCCA GAAATCGTTG CACATTCGCA TTACACTGTA CAAGGCCAAG	10620
	ACTTCCCAGC AGGTAATGGT TCTAGTGCAT CAGATTACTT TAAGTTATCT AATGGTAGTG	10680
	ACATTGCAGA TGCAACTATT ACATGGGTAA GTGGACAAGC GCCAAATAAA GATAATACAC	10740
15	GTATTGGTGA AGATATAACT GTAACGTCAC ATATCTTAAT TGATGGCGAA ACAACGCCGA	10800
	TTACGAAAAC AGCAACATAA AAAGTAGTAA GAACTGTACC GAAACATGTC TTTGAAACAG	10860
20	CCAGAGGTGT TTTATACCCA GGTGTTTCAG ATATGTATGA TGCGAAACAA TATGTTAAGC	10920
	CAGTAAATAA TTCTTGGTCG ACAAATGCGC AACATATGAA TTTCCAATTT GTTGAACAT	10980
	ATGGTCCTAA CAAAGATGTT GTAGGCATAT CTACTCGTCT TATTAGAGTG ACATATGATA	11040
25	ATAGACAAAC AGAAGATTTA ACTATTTTAT CTAAAGTTAA ACCTGACCCA CCTAGAATTG	11100
	ACGCAAACTC TGTGACATAT AAAGCAGGTC TTACAAACCA AGAAATTAAA GTTAATAACG	11160
	TATTAAATAA CTCGTCAGTA AAATTATTTA AAGCAGATAA TACACCATTA AATGTCACAA	11220
30	ATATTACTCA TGGTAGCGGT TTTAGTTCGG TTGTGACAGT AAGTGACGCG TTACCAAATG	11280
	GCGGAATTAA AGCAAAATCT TCAATTTCAA TGAACAATGT GACGTATACG ACGCAAGACG	11340
	AACATGGTCA AGTTGTTACA GTAACAAGAA ATGAATCTGT TGATTCAAAT GACAGTGCAa	11400
35	CAGTAACAGT GACACCACAA TTACAAGCAA CTACTGAAGG CGCTGTATTT ATTAAAGGTG	11460
	GCGA <sup>6</sup> GGTTT TGATTTGGA CACGTAGAAA GATTTATTCA AAACCCGCCA CATGGGGCAA	11520
	CGGTTGCATG GCATGATAGT CCAGATACAT GGAAGAATAC AGTCGGTAAC ACTCATAAAA	11580
40	CTGCGGTTGT AACATTACCT AATGGTCAAG GTACGCGTAA TGTGAAGTT CCAGTCAAAG	11640
	TTTATCCAGT TGCTAATGCA AAGGCGCCAT CACGTGATGT GAAAGGTCAA AATTTGACTA	11700
45	ATGGAACGGA TGCGATGAAC TACATTACAT TTGATCCAAA TACAAACACA AATGGTATCA	11760
	CTGCAGCATG GGCAATAGA CAACAACCAA ATAACCAACA AGCAGGCGTG CAACATTTAA	11820
	ATGTCGATGT CACATATCCA GGTATTTTAC CTGCTAAACG AGTTCCTGTT ACTGTTAATG	11880
50	TATATCAATT TGAATTCCT CAAACTACTT ATACGACAAC GGTGGAGGC ACTTTAGCAA	11940
	GTGGTACGCA AGCATCAGGA TATGCACATA TGCAAAATGC TACTGGTTTA CCAACAGATG	12000

55



	TGAATAAACCC GAATGTGGCT AAAGTCGTTA ACGCAAAATA TGACGTCATC TATAACGGAC	12120
	ATACTTTTGC AACATCTTTA CCAGCGAAAT TTGTAGTAAA AGATGTGCAA CCAGCGAAAC	12180
5	CAACTGTGAC TGAAACAGCG GCAGGAGCGA TTACAATTGC ACCTGGAGCA AACCAAACAG	12240
	TGAATACACA TGCCGGTAAC GTAACGACAT ACGCTGATAA ATTAGTTATT AAACGTAATG	12300
	GTAACGTTGT GACGACATTT ACACGTCGCA ATAATACGAG TCCATGGGTG AAAGAAGCAT	12360
10	CTGCAGCAAC TGTAGCAGGT ATTGCTGGAA CTAATAATGG TATTACTGTT GCAGCAGGTA	12420
	CTTTCAACCC TGCTGATACA ATTCAAGTTG TTGCAACCGA AGGAAGCGGA GAGACAGTGA	12480
15	GTGATGAGCA ACGTAGTGAT GATTTTCACAG TTGTCGCACC ACAACCGAAC CAAGCGACTA	12540
	CTAAGATTGT GCAAAATGGT CATATTGATA TCACGCCTAA TAATCCATCA GGACATTTAA	12600
	TTAATCCAAC TCAAGCAATG GATATTGCTT ACACTGAAAA AGTGGGTAAT GGTGCAGAAC	12660
20	ATAGTAAGAC AATTAAATGTT GTTCGTGGTC AAAATAATCA ATGGACAATT GCGAATAAGC	12720
	CTGACTATGT AACGTTAGAT GCACAAACTG GTAAAGTGAC GTTCAATGCC AATACTATAA	12780
	AACCAAATTC ATCAATCACA ATTACTCCGA AAGCAGGTAC AGGTCACTCA GTAAGTAGTA	12840
25	ATCCAAGTAC ATTAAGTACA CCGGCAGCTC ATACTGTCAA CACAACTGAA ATTGTGAAAG	12900
	ATTATGGTTC AAATGTAACA GCAGCTGAAA TTAACAATGC AGTTCaAGTT GCTAATAAAC	12960
	GTACTGCAAC GATTAAAAAT GGCACAGCAA TGCCTACTAA TTTAGCTGGT GGTAGCACAA	13020
30	CGACGATTCC TGTGACAGTA ACTTACAATG ATGGTAGTAC TGAAGAAGTA CAAGAGTCCA	13080
	TTTTTCACAAA AGCGGATAAA CGTGAGTTAA TCACAGCTAA AAATCATTTA GATGATCCAG	13140
	TAAGCACTGA AGGTAAAAAG CCAGGTACAA TTACGCAGTA CAATAATGCA ATGCATAATG	13200
35	CGCAACAACA AATCAATACT GCGAAAACAG AAGCACAACA AGTGATTAAT AATGAGCGTG	13260
	CAACACCACA ACAAGTTTCT GACGCACTAA CTAAAGTTCTG TGCAGCACAA ACTAAGATTG	13320
	ATCAAGCTAA AGCATTACTT CAAAATAAAG AAGATAATAG CCAATTAGTA ACGTCTAAAA	13380
40	ATAACTTACA AAGTTCTGTG AACCAAGTAC CATCAACTGC TGGTATGACG CAACAAAGTA	13440
	TTGATAACTA TAATGCGAAG AAGCGTGAAG CAGAACTGA AATAACTGCA GCTCAACGTG	13500
45	TTATTGACAA TGGCGATGCA ACTGCACAAC AAATTCAGA TGAAAAACAT CGTGTCGATA	13560
	ACGCATTAAC AGCATTAAC CAAGCGAAAC ATGATTTAAC TGCAGATACA CATGCCTTAG	13620
	AGCAAGCAGT GCAACAATTG AATCGCACAG GTACAACGAC TGGAAGAAG CCGGCAAGTA	13680
50	TTACTGCTTA CAATAATTCG ATTCGTGCAC TTCAAAGTGA CTTAACAAGT GCTAAAAATA	13740
	GCGCTAATGC TATTATTCAA AAGCCAATAA GAACAGTACA AGAAGTGCAA TCTGCGTTAA	13800
55		

	CTGATAATAG	TGCTTTAAAA	ACTGCTAAGA	CGAAACTTGA	TGAAGAAATC	AATAAATCAG	13920
	TAACTACTGA	TGGTATGACA	CAATCATCAA	TCCAAGCATA	TGAAAATGCT	AAACGTGCGG	13980
5	GTCAAACAGA	ATCAACAAAT	GCACAAAATG	TTATTAACAA	TGGTGATGCG	ACTGACCAAC	14040
	AAATTGCCGC	AGAAAAAACA	AAAGTAGAAG	AAAAATATAA	TAGCTTAAAA	CAAGCAATTG	14100
	CTGGATTAAC	TCCAGACTTG	GCACCATTAC	AAACTGCAAA	AACTCAGTTG	CAAAATGATA	14160
10	TTGATCAGCC	AACGAGTACG	ACTGGTATGA	CAAGCGCATC	TATTGCAGCA	TTTAATGAAA	14220
	AACTTTCAGC	AGCTAGAACT	AAAATTCAAG	AAATTGATCG	TGTATTAGCC	TCACATCCAG	14280
	ATGTTGCGAC	AATACGTCAA	AACGTGACAG	CAGCGAATGC	CGCTAAATCA	GCACTTGATC	14340
15	AAGCACGTAA	TGGCTTAACA	GTCGATAAAG	CGCCTTTAGA	AAATGCGAAA	AATCAACTAC	14400
	AACATAGTAT	TGACACGCAA	ACAAGTACAA	CTGGTATGAC	ACAAGACTCT	ATAAATGCAT	14460
20	ACAATGCGAA	GTTAACAGCT	GCACGTAATA	AGATTCAACA	AATCAATCAA	GTATTAGCAG	14520
	GTTCAACGAC	TGTAGAACAA	ATTAATACAA	ATACGTCTAC	AGCAAATCAA	GCTAAATCTG	14580
	ATTTAGATCA	TGCACGTCAA	GCTTTAACAC	CAGATAAAGC	GCCGCTTCAA	ACTGCGAAAA	14640
25	CGCAATTAGA	ACAAAGCATT	AATCAACCAA	CGGATACAAC	AGGTATGACG	ACCGCTTCGT	14700
	TAAATGCGTA	CAACCAAAAA	TTACAAGCAG	CGCGTCAAAA	GTAACTGAA	ATTAATCAAG	14760
	TGTTGAATGG	CAACCCAACT	GTCCAAAATA	TCAATGATAA	AGTGACAGAG	GCAAACCAAG	14820
30	CTAAGGATCA	ATTAAATACA	GCACGTCAAG	GTTTAACATT	AGATAGACAG	CCAGCGTTAA	14880
	CAACATTACA	TGGTGCATCT	AACTTAAACC	AAGCACAACA	AAATAATTTC	ACGCAACAAA	14940
	TTAATGCTGC	TCAAAATCAT	GCTGCGCTTG	AAACAATTAA	GTCTAACATT	ACGGCTTTAA	15000
35	ATACTGCGAT	GACGAAATTA	AAAGACAGTG	TTGCGGATAA	TAATACAATT	AAATCAGATC	15060
	AAAATTACAC	TGACGCAACA	CCAGCTAATA	AACAAGCGTA	TGATAATGCA	GTTAATGCGG	15120
40	CTAAAGGTGT	CATTGGAGAA	ACGACTAATC	CAACGATGGA	TGTTAACACA	GTGAACCAAA	15180
	AAGCAGCATC	TGTTAAATCG	ACGAAAGATG	CTTTAGATGG	TCAACAAAAC	TTACAACGTG	15240
	CGAAAACAGA	AGCAACAAAT	GCGATTACGC	ATGCAAGTGA	TTTAAACCAA	GCACAAAAGA	15300
45	ATGCATTAAC	ACAACAAGTG	AATAGTGcAC	AAAACGTGCA	AGCAGTAAAT	GATATTAAAC	15360
	AAACGACTCA	AAGCTTAAAT	ACTGCTATGA	CAGGTTTAAA	ACGTGGCGTT	GCTAATCATA	15420
	ACCAAGTCGT	ACAAAGTGAT	AATTATGTCA	ACGCAGATAC	TAATAAGAAA	AATGATTACA	15480
50	ACAATGCATA	CAACCATGCG	AATGACATT	TTAATGGTAA	TGCACAACAT	CCAGTTATAA	15540
	CACCAAGTGA	TGTTAACAAT	GCTTTATCAA	ATGTCACAAG	TAAAGAACAT	GCATTGAATG	15600
55							

EP 0 786 519 A2

	ATTTAAATAA TGCACAACGT CAAAACCTTAC AATCGCAAAT TAATGGTGCG CATCAAATTG	15720
	ATGCAGTTAA TACAATTAAG CAAAATGCAA CAACTTGAA TAGTGCAATG GGTAACCTAA	15780
5	GACAAGCTGT TGCAGATAAA GATCAAGTGA AACGTACAGA AGATTATGCG GATGCAGATA	15840
	CAGCTAAACA AAATGCATAT AACAGTGCAG TTTCAAGTGC CGAAACAATC ATTAATCAAA	15900
	CAACAAATCC AACCATGTCT GTTGATGATG TTAATCGTGC AACTTCAGCT GTTACTTCTA	15960
10	ATAAAAATGC ATTAAATGGT TATGAAAAAT TAGCACAATC TAAACAGAT GCTGCAAGAG	16020
	CAATTGATGC ATTACCACAT TTAAATAATG CACAAAAAGC AGATGTTAAA TCTAAAATTA	16080
	ATGCTGCATC AAATATTGCT GCGGTAAATA CTGTTAAACA ACAAGGTACA GATTTAAATA	16140
15	CAKCGATGGg TAACTTGCAA GGTGCAATCA ATGATGAACA AACGACGCTT AATAGTCAAA	16200
	ACTATCAAGA TCGACACCT AGTAAGAAAA CAGCATACAC AAATGCGGTA CAAGCTGCGA	16260
20	AAGATATTTT AAATAAATCA AATGGTCAAA ATAAAACGAA AGATCAAGTT ACTGAAGCGA	16320
	TGAATCAAGT GAATTCTGCT AAAAATAACT TAGATGGTAC GCGTTTATTA GATCAAGCGA	16380
	nCAAaCAGCA AAACAGCAGT TAAATAATAT GACGCATTTA ACAACTGCAC AAAAAACGAA	16440
25	TTTAACAAAC CAAATTAATA GTGGTACTAC TGTCGCTGGT GTTCAAACGG TTCAATCAAA	16500
	TGCCAATACA TTAGATCAAG CCATGAATAC GTTAAGACAA AGTATTGCCA ACAAAGATGC	16560
	GACTAAAGCA AGTGAAGATT ACGTAGATGC TAATAATGAT AAGCAAACAG CATATAACAA	16620
30	CGCAGTAGCT GCTGCTGAAA CGATTATTAA TGCTAATAGT AATCCAGAAA TGAATCCAAG	16680
	TACGATTACA CAAAAGCAG AGCAAGTGAA TAGTTCTAAA ACGGCACTTA ACGGTGATGA	16740
	AACTTAGCT GCTGCAAAAC AAAATGCGAA AACGTACTTA AACACATTGA CAAGTATTAC	16800
35	AGATGCTCAA AAGAACAATT TGATTAGTCA AATTACTAGT GCGACAAGAG TGAGTGGTGT	16860
	TGATACTGTA AAACAAAATG CGCAACATCT AGACCAAGCT ATGGCTAGCT TACAGAATGG	16920
	TATTAACAAC GAATCTCAAG TGAAATCATC TGAGAAATAT CGTGATGCTG ATACAAATAA	16980
40	ACAACAAGAG TATGATAATG CTATTACTGC AGCGAAAGCG ATTTTAAATA AATCGACAGG	17040
	TCCAAACACT GCGCAAAATG CAGTTGAAGC AGCATTACAA CGTGTTAATA ATGCGAAAGA	17100
45	TGCATTGAAT GGTGATGCAA AATTAATTGC AGCTCAAAAC GCAGCGAAAC AACATTTAGG	17160
	TACTTTAACG CATATCACTA CAGCTCAACG TAATGATTTA ACAAATCAAA TTTCACAAGC	17220
	TACAAACTTA GCTGGTGTG AATCTGTAA ACAAATGCG AATAGTTTAG ATGGTGCTAT	17280
50	GGGTAACCTA CAAACGGCTA TCAACGATAA GTCAGGAACA TTAGCGAGCC AAAACTTCTT	17340
	GGATGCTGAT GAGCAAAAAC GTAATGCATA CAATCAAGCT GTATCAGCAG CCGAAACCAT	17400
55		

	TGTTAATAAT GCGAAACATG CATTAAATGG TACGCAAAAC TTAAACAATG CGAAACAAGC	17520
	ACCGATTACA GCAATCAATG GCGCATCTGA TTTAAATCAA AAACAAAAAG ATGCATTAAA	17530
5	AGCACAAGCT AATGGTGCTC AACGCGTATC TAATGCACAA GATGTACAGC ACAATGCGAC	17640
	TGAACTGAAC ACGGCAATGG GCACATTAAA ACATGCCATC GCAGATAAGA CGAATACGTT	17700
	AGCAAGCAGT AAATATGTTA ATGCCGATAG CACTAAACAA AATGCTTACA CAACTAAAGT	17760
10	TACCAATGCT GAACATATTA TTAGCGGTAC GCCAACGGTT GTTACGACAC CTTCAGAAGT	17820
	AACAGCTGCA GCTAATCAAG TAAACAGCGC GAAACAAGAA TTAAATGGTG ACGAAAGATT	17880
	ACGTGAAGCA AAACAAAACG CCAATACTGC TATTGATGCA TTAACACAAT TAAATACACC	17940
15	TCAAAAAGCT AAATTAAAAG AACAGTGGG ACAAGCCAAT AGATTAGAAG ACGTACAAAC	18000
	TGTTCAAACA AATGGACAAG CATTGAACAA TGCAATGAAA GGCTTAAGAG ATAGTATTGC	18060
20	TAACGAAACA ACAGTCAAAA CAAGTCAAAA CTATACAGAC GCAAGTCCGA ATAACCAATC	18120
	AACATATAAT AGCGCTGTGT CAAATGCGAA AGGTATCATT AATCAAATA ACAATCCGAC	18180
	TATGGATACT AGTGCGATTA CCCAAGCTAC AACACAAGTG AATAATGCTA AAAATGGTTT	18240
25	AAACGGTGCT GAAAACCTAA GAAATGCACA AAACACTGCT AAGCAAACT TAAATACATT	18300
	ATCACACTTA ACAAATAACC AAAAATCTGC CATCTCATCA CAAATTGATC GTGCAGGTCA	18360
	TGTGAGTGAG GTAACCTGTA CTAAAAATGC AGCAACTGAG TTGAATACGC AAATGGGTAA	18420
30	CTTGGAACAA GCTATCCATG ATCAAAACAC AGTTAAACAA AGTGTTAAAT TTAATGATGC	18480
	AGATAAAGCT AAACGTGATG CGTATACAAA TGCGGTAAGC AGAGCTGAAG CAATTCTGAA	18540
	TAAAACGCAA GGTGCAAATA CGTCTAAACA AGATGTTGAA GCGGCTATTC AAAATGTTTC	18600
35	AACTGCTAAA AATGCATTGA ATGGTGATCA AAACGTTACA AATGCGAAGA ATGCAGCTAA	18660
	AAATGCATTA AATAACTTAA CGTCAATTAA TAATGCACAA AAACGTGACT TAACAATAA	18720
	AATTGATCAA GCAACAACCTG TAGCTGGTGT TGAAGCTGTA TCTAATACGA GTACACAATT	18780
40	GAATACAGCG ATGGCTAACT TGCAAAATGG TATTAATGAT AAAACAAATA CACTAGCAAG	18840
	TGAAAACCTAT CATGATGCTG ATTCAGATAA GAAAACCTGCT TATACTCAAG CCGTTACGAA	18900
45	CGCAGAAAAT ATTTTAAATA AAAATAGTGG ATCAAATTTA GACAAAACCTG CCGTTGAAAA	18960
	CGCGTTGTCA CAAGTTGCTA ATGCGAAAGG TGCCCTAAAT GGTAACCATA ATTTAGAGCA	19020
	AGCTAAATCA AATGCAAACA CTACTATAAA CGGACTTCAA CATTTAACAA CTGCTCAAAA	19080
50	AGATAAATTG AAACAACAAG TGCAACAAGC ACAAATGTT GCAGGTGTAG ATACTGTAA	19140
	ATCAAGTGCC AACACATTAA ATGGTGCTAT GGGTACGTTA AGAAATAGCA TACAAGATAA	19200

55

	TAACAATGCT GTTGATAGTG CTAATGGTGT CATTAAATGCA ACAAGCAATC CAAATATGGA	19320
	TGCTAATGCA ATTAACCAAA TCGCTACACA AGTGACATCA ACGAAAAATG CATTAGATGG	19380
5	TACACATAAT TTAACGCAAG CGAAACAAAC AGCAACAAAT GCCATCGATG GTGCTACTAA	19440
	CTTAAATAAA GCGCAAAAAG ATGCGTTAAA AGCACAAAGTT ACAAGTGCGC AACGTGTTGC	19500
	AAATGTAACA AGTATCCAAC AAAGTGC AAACTGCAAA TGAAGTTAAT ACAGCTATGG GTCAATTACA	19560
10	ACATGGTATT GATGATGAAA ATGCAACAAA ACAAACTCAA AAATATCGTG ACGcTGAACA	19620
	AAGTAAGAAA ACTGCTTATG ATCAAGCTGT AGCTGCTGCG AAAGCAATTT TAAATAAACA	19680
	AACAGGTTCA AATTCAGATA AAGCAGCAGT TGACCGTGCA TTACAACAAG TAACAAGTAC	19740
15	GAAAGATGCA TTGAATGGTG ATGCAAAACT GGCAGAAGCG AAAGCGGCAG CTAAACAAAA	19800
	CTTAGGCACT TTAAACCATA TTACGAATGC ACAACGTACT GACTTAGAAG GCCAAATCAA	19860
20	TCAAGCGACG ACTGTTGATG GCGTTAATAC TGTA AAAACA AATGCCAATA CATTAGACGG	19920
	CGCAATGAAT AGCTTACAAG GTTCAATCAA TGATAAAGAT GCGACATTAA GAAATCAAAA	19980
	TTATCTTGAT GCGGATGAAT CAAAACGAAA TGCATATACG CAAGCTGTCA CAGCGGCTGA	20040
25	AGGCATTTTA AATAAACAAA CTGGTGGTAA CACATCTAAA GCAGACGTTG ATAATGCATT	20100
	AAATGCAGTT ACAAGAGCGA AAGcGgCTTT AAATGGTGCT GACAACTTAA GAAATGCGAA	20160
	AACTTCAGCA ACAAATACGA TTGATGGTTT ACCTAACTTA ACACAATTAC AAAAAGACAA	20220
30	CTTGAAGCAT CAAGTTGAaC AAGCGCAAAA TGTAGCAGGT GTAAATGGTG TTAAAGATAA	20280
	AGGTAATACG TTAAATACTG CCATGGGTGC ATTACGTACA AGTATCCAAA ATGATAATAC	20340
	GACGAAAACA AGTCAAAATT ATCTTGATGC ATCTGACAGC AACAAAAATA ATTACAATAC	20400
35	TGCTGTAAAT AATGCAAAATG GTGTTATTAA TGCAACGAAC AATCCAAATA TGGATGCTAA	20460
	TGCGATTAAAT GGCATGGCAA ATCAAGTCAA TACAACAAAA GCAGCGTTAA ATGGTGCACA	20520
	AAACTTAGCT CAAGCTAAAA CAAATGCGAC GAACACAATT AACACGCAC ATGACTTAAA	20580
40	CCAAAAACAA AAAGATGCAT TAAAAACACA AGTTAACAAT GCACAACGTG TATcTGATGC	20640
	AAATAACGTT CAACACACTG CAACTGAATT GAACAGTGCG ATGACAGCAC TTAAAGCAGC	20700
45	TATTGCTGAT AAAGAAAGAA CAAAAGCAAG CGGTAATTAT GTCAATGCTG ATCAAGAAAA	20760
	ACGTCAAGCG TATGATTCAA AAGTGACTAA CGCTGAAAAT ATCATTAGTG GTACACCGAA	20820
	TGCGACATTA ACAGTCAATG ACGTAAATAG TGCGGCATCA CAAGTCAATG CGGCTAAAAC	20880
50	AGCATTAAAT GGTGATAACA ACTTACGTGT AGCGAAAGAG CATGCCAACA ATACAATTGA	20940
	CGGCTTAGCA CAATTGAATA ATGCACAAAA AGCAAAATTA AAAGAACAAG TTCAAAGTGC	21000

55

GAAAGGCTTA AGAGATAGTA TTGCGAATGA AGCAACAATT AAAGCAGGTC AAAACTACAC 21120  
 TGACGCAAGT CCAAATAATC GTAACGAGTA CGACAGTGCA GTTACTGCAG CAAAAGCAAT 21180  
 5 CATTAAATCAA ACATCGAACC CAACGATGGA ACCAAATACT ATTACGCAAG TAACATCACA 21240  
 AGTGACAACCT AAAGAACAGG CATTAAATGG TGCGCGAAAC TTAGCTCAAG CTAAGACAAC 21300  
 TGCGAAAAAC AACTTGAATA ACTTAACATC AATTAACAAT GCACAAAAAG ATGCGTTAAC 21360  
 10 GCGTAGcATT GATGGTGCAA CAACAGTAGC TGGTGTAAT CAAGAACTG CAAAAGCAAC 21420  
 AGAATTAAAT AACGCAATGC ATAGTTTACA AAATGGTATC AATGATGAGA CACAAACAAA 21480  
 ACAAACCTCAG AAATACCTAG ATGCAGAGCC AAGTAAGAAA TCAGCTTATG ATCAAGCAGT 21540  
 15 AAATGCAGCG AAAGCAATTT TAACAAAAGC TAGTGGTCAA AATGTAGACA AAGCAGCAGT 21600  
 TGAACAAGCA TTGCAAAATG TGAACAGTAC GAAGACGGCG TTGAACGGTG ATGCGAAATT 21660  
 AAATGAAGCT AAAGCAGCTG CGAAACAAAC GTTAGGTACA TTAACACACA TTAATAATGC 21720  
 20 ACAACGTACA GCGTTAGACA ATGAAATTAC ACAAGCAACA AATGTTGAAG GTGTTAATAC 21780  
 AGTTAAAGCC AAAGCGCAAC AATTAGATGG TGCTATGGGT CAATTAGAAA CATCAATTCTG 21840  
 25 TGATAAAGAC ACGACGTTAC AAAGTCAAAA TTATCAAGAT GCTGATGATG CTAAACGAAC 21900  
 TGCTTATTCT CAAGCAGTAA ATGCAGCAGC AACTATTTTA AATAAAACag CTGGCGGTAA 21960  
 TACACCTAAA GCAGATGTTG AAAGAGCAAT GCAAGCTGTT ACACAAGCAA ATACTGcATT 22020  
 30 AAACGGTATT CAmAACITAG ATCGTGCGAA ACaRGCTGCT AACACAGCGA TTACAAATGC 22080  
 TTCGGACTTA AATACAAAAC mAAAAGAAGC ATTAaAgCA CAAGTAACAA GTGCAGGACG 22140  
 TGTATCTGCA GCAAATGGTG TTGAACATAC TGCGACTGAA TTAAATACTG CGATGACAGC 22200  
 35 TTTAAAGCGT GCCATTGCTG ATAAAGCTGA GACAAAAGCT AGTGGTAACT ATGTCAATGC 22260  
 TGATGCGAAT AAACGTCAAG CATATGATGA AAAAGTTACA GCTGCCGAAA ATATCGTTAG 22320  
 TGGTACACCA ACACCAACGT TAACACCAGC AGATGTTACA AATGCAGCAA CGCAAGTAAC 22380  
 40 GAATGCTAAG ACGCAGTTAA ACGGTAATCA TAATTTAGAA GTAGCGAAAC AAAATGCTAA 22440  
 CACTGCAATT GATGGTTTAA CTTCTTTAAA TGGTCCGCAA AAAGCAAAAC TTAAAGAACA 22500  
 AGTGGGTCAA GCGACGACGT TGCCAAATGT TCAAAGTGT CGTGATAATG CACAAACATT 22560  
 45 AAACACTGCA ATGAAAGGTC TACGAGATAG CATTGCGAAT GAAGCAACGA TTAAAGCAGG 22620  
 TCAAACTAC ACAGATGCAA GTCAAAACAA ACAAAGTAC TACAACAGTG CAGTCACTGC 22680  
 50 AGCAAAAGCA ATCATTGGTC AAACAAC TAGTCCGCAAT AATGCGCAAG AAATTAATCA 22740  
 AGCGAAAGAC CAAGTGACAG CTAAACAACA AGCGTTAAAC GGTCAAGAAA ACTTAAGAAC 22800

55

	AGATGCAGTG AAACGTCAAA TCGAAGGTGC AACGCATGTT AATGAAGTAA CACAAGCACA	22920
	AAATAATGCG GATGCaTTAA ATACAGCTAT GACGAAGTTG AAAAAATGGTA TTCAAGATCA	22980
5	GAATACGATT AAGCAAGGTG TTAACCTCAC TGATGCCGAC GAAGCGAAAC GTAATGCATA	23040
	TACAAATGCA GTGACGCAAG CTGAACAAAT TTAAATAAA GCACAAGGTC CAAATACTTC	23100
	AAAAGACGGT GTCGAAACTG CGTTAGaAA TGTACAACGT GCTAAAAACG AATTGAACGG	23160
10	TAATCAAAAT GTTGCGAAG CTAAGACAAC TGCGAAAAAT GCATTGAATA ACCTAACATC	23220
	AATTAATAAT GCACAAAAAG AAGCATTGAA ATCACAATT GAAGGTGCGA CAACAGTTGC	23280
15	AGGTGTAAAT CAAGTGTCTA CAACGGCATC TGAATTAAAT ACAGCAATGA GCAACTTACA	23340
	AAATGGTATT AATGATGAAG CAGCTACAAA AGCAGCGCTT AATGGTACTC AAAACCTTGA	23400
	AAAAGCTAAA CAACACGCAA ATACAGCAAT TGACGGTTTA AGCCATTTAA CAAATGCACA	23460
20	AAAAGAGGCA TTAACAAT TGGTACAACA ATCGACTACT GTTGCGAAG CACAAGGTAA	23520
	TGAGCAAAAA GCAACAATG TTGATGCAGC AATGGACAAA TTACGTCAA GTATTGCAGA	23580
	TAATGCGACA ACAAACAAA ACCAAAATTA TACTGATGCA AGTCAGAATA AAAAGGATGC	23640
25	GTACAATAAT GCTGTCACAA CTGCACAAGG TATTATTGAT CAAACTACAA GTCCAACTTT	23700
	AGATCCGACT GTTATCAATC AAGCTGCTGG ACAAGTAAGC ACAACTAAAA ATGCATTAAA	23760
	TGGTAATGAA AACCTAGAGG CAGCGAAACA ACAAGCGTCA CAATCATTAG GTTCATTAGA	23820
30	TAACTTAAAT AATGCGCAAA AACAAACAGT TACTGATCAA ATTAATGGCG CGCATACTGT	23880
	TGATGAAGCA AATCAAATTA AGCAAAATGC GCAAACTTA AATACAGCGA TGGGTAACTT	23940
	GAAACAAGCG ATAGcTGACA AAGATGCTAC GAAAGCGACA GTTAACCTCA CTGATGCAGA	24000
35	TCAAGCAAAA CAACAAGCAT ATAACaCTGC TGTTACAAAT GCTGAAAATA TCATTTCAAA	24060
	AGCTAATGGC GGCAATGCAA CACAAGCTGA AGTTGAACAA GCAATCAAAC AAGTTAATGC	24120
	TGCAAAACAA GCATTAAATG GTAATGCCAA CGTTCAACAT GCAAAAGACG AAGCAACAGC	24180
40	ATTAATTAAT AGCTCTAATG ACCTTAACCA AGCACAAAAA GACGCATTAA AACAAACAGT	24240
	TCAAAATGCA ACTACTGTAG CTGGTGTAAT CAATGTTAAA CAAACAGCAC AAGAGTTAAA	24300
45	CAATGCTATG ACACAATTAA AACAAAGGCAT TGCAGATAAA GAACAAACAA AAGCTGATGG	24360
	TAACTTTGTC AATGCAGATC CTGATAAGCA AAATGCATAT AATCAAGCAG TAGCGAAAGC	24420
	TGAAGCATT AATTAGTGetA CGCCTGATGT TGTCGTTACA CCTAGCGAAA TTAAGTCAGC	24480
50	GTAAATAAAA GTTACGCAAG CTAAAAATGA TTAAATGGT AATACAAACT TAGCAACGGC	24540
	GAAACAAAAT GTTCAACATG CTATTGATCA ATTGCCAAAC TTAAACCAAG CGCAACGTGA	24600

55

## EP 0 786 519 A2

	AGCGGCGACA ACGCTTAATG ACGCGATGAC ACAATTGAAA CAAGGTATTG CGAATAAAGC	24720
	ACAAATTAAA GGTAGCGAGA ACTATCACGA TGCTGATACT GACAAGCAAA CAGCATATGA	24780
5	TAATGCAGTA ACAAAGCAG AAGAATTGTT AAAACAAACA ACAAATCCAA CAATGGATCC	24840
	AAATACAATT CAACAAGCAT TAACTAAAGT GAATGACACA AATCAAGCAC TTAACGGTAA	24900
	TCAAAAATTA GCTGATGCCA AACAAGATGC TAAGACAACA CTTGGTACAC TAGATCATTT	24960
10	AAATGATGCT CAAAAACAAG CGCTAACAAAC TCAAGTTGAA CAAGCACCAG ATATTGCAAC	25020
	AGTTAATAAT GTTAAGCAAA ATGCTCAAAA TCTGAATAAT GCTATGACTA ACTTAAACAA	25080
	TGCATTACAA GATAAACTG AGACATTAAA TAGCATTAAAC TTTACTGATG CAGATCAAGC	25140
15	TAAGAAAGAT GCTTATACTA ATGCGGTTTC ACATGCAGAA GGTATTTTAT CTAAAGCAAA	25200
	TGGCAGCAAT GCAAGTCAAA CTGAAGTGGA ACAAGCGATG CAACGTGTGA ACGAAGCGAA	25260
20	ACAAGCATTG AATGGTAATG ACAATGTACA ACGTGCAAAA GATGCAGCGA AACAAGTGAT	25320
	TACAAATGCA AATGATTTAA ATCAAGCAAT GACACAATTG AAACAAGGTA TTGCAGATAA	25380
	AGACCAAACCT AAAGCAAATG GTAACCTTGT CAATGCTGAT ACTGATAAGC AAAATGCTTA	25440
25	CAACAATGCG GTAGCACATG CTGAACAAAT AATTAGTGGT ACACCAAATG CAAACGTGGA	25500
	TCCACAACAA GTGGCTCAAG CGTTACAACA AGTGAATCaA GCTAAGGGTG ATTTAAACGG	25560
	TAACCATAAC TTACAAGTTG CTAAAGACAA TGCAAATACA GCCATTGATC AGTTACCAAA	25620
30	CTTAAATCAA CCACAAAAAA CAGCATTAAA AGACCAAGTG TCGCATGCAG AACTTGTTAC	25680
	AGGTGTTAAT GCTATTAAGC AAAATGCTGA TGCGTTAAAT AATGCaATGG GTACATTGAA	25740
	ACAACAAATT CAAGCGAACA GTCAAGTACC ACAGTCAGTT GACTTTACAC AAGCGGATCA	25800
35	AGACAAACAA CAAGCATATA ACAATGCGGC TAACCAAGCG CAACAAATCG CAAATGGCAT	25860
	ACCAACACCT GTATTGACGC CTGATACAGT AACACAAGCA GTGACAACCTA TGAATCAAGC	25920
	GAAAGATGCA TTAAACGGTG ATGAAAAATT AGCACAAGCG AAACAAGAAG CTTTAGCAAA	25980
40	TCTTGATACG TTACGCGATT TAAATCAACC ACAACGTGAT GCATTACGTA ACCAAATCAA	26040
	TCAAGCACAA GCGTTAGCTA CAGTTGAACA AACTAAACAA AATGCACAAA ATGTGAATAC	26100
45	aGCaATGAGT AACTTGAAAC aAGGTATTGC aAACAAAGAT ACTGTCAAAG CAAGTGAGAA	26160
	CTATCATGAT GCTGATGCCG ATAAGCAAAC AGCATATACA AATGCAGTGT CTCAAGCGGA	26220
	AGGTATTATC AATCAAACGA CAAATCCAAC GCTTAACCCA GATGAAATAA CACGTGCATT	26280
50	AACTCAAGTG ACTGATGCTA AAAATGGCTT AAACGGTGAA GCTAAATTGG CAACTGAAAA	26340
	GCAAAATGCT AAAGATGCCG TAAGTGGGAT GACGCATTTA AACGATGCTC AAAACAAGC	26400

55



	AGCAACGAGC CTAGATCAAG CAATGGATCA ATTATCACAA GCTATTAATG ATAAAGCTCA	26520
	AACATTAGCG GACGGTAATT ACTTAAATGC AGATCCTGAC AAACAAAATG CGTATAAACA	26580
5	GGCAGTAGCA AAAGCTGAAG CATTATTGAA TAAACAAAGT GGTACTAATG AAGTACAAGC	26640
	ACAAGTTGAA AGCATCACTA ATGAAGTGAA CGCAGCGAAA CAAGCATTAA ATGGTAATGA	26700
	CAATTTGGCA AATGCAAAAC AACAAAGCAA ACAACAATTG GCGAACTTAA CACACTTAAA	26760
10	TGATGCACAA AAACAATCAT TTGAAAGTCA AATTACACAA GCGCCACTTG TTACAGATGT	26820
	CACTACGATT AATCAAAAAG CACAAACGTT AGATCATGCG ATGGAATTAT TAAGAAATAG	26880
	TGTTGCGGAT AATCAAACGA CATTAGCGTC TGAAGATTAT CATGATGCAA CTGCGCAAAG	26940
15	ACAAAATGAC TATAACCAAG CTGTAACAGC TGCTAATAAT ATAATTAATC AAACCTACATC	27000
	GCCTACGATG AATCCAGATG ATGTTAATGG TGCAACGACA CAAGTGAATA ATACGAAAGT	27060
20	TGCATTAGAT GGTGATGAAA ACCTTGCGAGC AGCTAAACAA CAAGCAAACA ACAGACTTGA	27120
	TCAATTAGAT CATTTGAATA ATGCGCAAAA GCAACAGTTA CAATCACAAA TTACGCAATC	27180
	ATCTGATATT GCTGCAGTTA ATGGTCACAA ACAACAGCA GAATCTTTAA ATACTGCGAT	27240
25	GGGTAACCTTA ATTAATGCGA TTGCAGATCA TCAAGCCGTT GAACAACGTG GTAACCTCAT	27300
	CAATGCTGAT ACTGATAAAC AAACCTGCTTA TAATACAGCG GTAAATGAAG CAGCAGCAAT	27360
	GATTAACAAA CAAACTGGTC AAAATGCGAA CCAAACAGAA GTAGAACAAG CTATTACTAA	27420
30	AGTTCAAACA ACACTTCAAG CGTTAAATGG AGACCATAAT TTACAAGTTG CTAACAAACA	27480
	TGCGACGCAA GCAATTGATG CTTTAACAAG CTTAAATGAT CCTCAAAAAA CAGCATTAAA	27540
	AGACCAAGTT ACAGCTGCAA CTTTAGTAAC TGCAGTTCAT CAAATTGAAC AAAATGCGAA	27600
35	TACGCTTAAC CAAGCAATGC ATGGTTTAAG ACAGAGCATT CAAGATAACG CAGCAACTAA	27660
	AGCAATAGC AAATATATCA ACGAAGATCA ACCAGAGCAA CAAAACCTATG ATCAAGCTGT	27720
40	TCAAGCCGCA AATAATATTA TCAATGAACA AACTGCAACA TTAGATAATA ATGCGATTAA	27780
	TCAAGCAGCG ACAACTGTGA ATACAACGAA AGCAGCATTG CATGGTGATG TGAAGTTACA	27840
	AAATGATAAA GATCATGCTA AGCAAACGGT TAGTCAATTA GCACATCTAA ACAATGCACA	27900
45	AAAACATATG GAAGATACGT TAATTGATAG TGAAACAACCT AGAACAGCAG TTAAGCAAGA	27960
	TTTGACTGAA GCACAAGCAT TAGATCAACT TATGGATGCA TTACAACAAA GTATTGCTGA	28020
	CAAAGATGCA ACACGTGCGA GCAGTGCATA TGTCAATGCA GAACCGAATA AAAACAATC	28080
50	CTATGATGAA GCAGTTCAAA ATGCTGAGTC TATCATTGCA GGATTAAATA ATCCAATAT	28140
	CAATAAAGGT AATGTATCAA GTGCGACTCA AGCAGTAATA TCATCTAAAA ATGCATTAGA	28200

55

	TCAATTAACA CCAGCTCAAC AACAAAGCGCT AGAAAATCAA ATTAATAATG CAACAACCTCG	28320
	TGATAAAGTG GCTGAAATCA TTGCACAAGC GCAAGCATtA AATGAAGCGA TGAAAGCATT	28380
5	AAAAGAAAGT ATTAAGGATC AACCACAAAC TGAAGCAAGT AGTAAATTTA TTAACGAGGA	28440
	TCAAGCGCAA AAAGATGCTT ATACGCAAGC AGTACAACAC GCGAAAGATT TGATTAACAA	28500
	AACAACCTGAT CCTACATTAG CTAAATCAAT CATTGATCAA GCGACACAGG CAGTGACAGA	28560
10	TGCTAAAAAC AATTTACATG GTGATCAAAA ACTAGCTCAA GATAAGCAAC GTGCAACAGA	28620
	AACGTTAAAT AACTTGTCTA ACTTGAATAC ACCACAACGT CAAGCACTTG AAAATCAAAT	28680
	TAATAATGCA GCAACTCGTG GCGAAGTAGC ACAAAAATTA ACTGAAGCAC AAGCACTTAA	28740
15	CCAAGCAATG GAAGCTTTAC GTAATAGCAT TCAAGATCAA CAGCAAACGG AAGCGGGTAG	28800
	CAAGTTTATC AATGAAGATA AaCCaCmAAA AGrTGCTTAC CAAGCAGCAG TTCAAAATGC	28860
20	AAAAGATTTA ATTAATCAAA CTAACAATCC AACGCTTGAT AAAGCACAAG TTGAACAATT	28920
	GACACAAGCT GTTAACCAAG CTAAAGATAA CCTACACGGT GATCAAAAAC TTGCAGACGA	28980
	TAAACAACAT GCGGTTACTG ATTTAAATCA ATTAAATGGT TTGAATAATC CGCAACGTCA	29040
25	AGCACTTGAA AGCCAAATAA ACAACGCAGC AACTCGTGGC GAAGTAGCAC AAAAATTAGC	29100
	TGAAGCAAAA GCGCTTGATC AAGCAATGCA AGCATTACGT AATAGTATTC AAGATCAACA	29160
	ACAAACAGAA TCTGGTAGCA AGTTTATCAA TGAAGATAAA CCGCAAAAAG ATGCTTACCA	29220
30	AGCAGCAGTT CAAAATGCAA AAGATTTAAT TAACCAAACA GGTAATCCAA CACTCGACAA	29280
	ATCACAAGTA GAACAATTGA CACAAGCAGT AACAACTGCA AAAGATAATC TACATGGTGA	29340
	TCAAAAACCTT GCTCGTGATC AACACAAGC AGTAACAACCT GTAAATGCAT TGCCAAAACCTT	29400
35	AAATCATGCA CAACAACAAG CATTAACCTGA TGCTATAAAT GCAGCGCCTA CAAGAACAGA	29460
	GGTTGACAA CATGTTCAAA CTGCTACTGA ACTTGATCAC GCGATGGAAA CATTGAAAAA	29520
	TAAAGTTGAT CAAGTGAATA CAGATAAGGC TCAACCAAAT TAACTGAAG CGTCAACTGA	29580
40	TAAAAAAGAA GCAGTAGATC AAGCGTTACA AGCTGCAGAA AGCATTACAG ATCCAACATA	29640
	TGGTTCAAAT GCGAATAAAG ACGCTGTAGA CCAAGTATTA ACTAAGCTTC AAGAAAAAGA	29700
45	AAATGAGTTA AATGGTAATG AGAGAGTCGC TGAAGCTAAA ACACAAGCGA AACAACTAT	29760
	TGACCAATTA ACACATTTAA ATGCTGATCA AATTGCAACT GCTAAACAAA ACATTGATCA	29820
	AGCGACGAAA CTTCAACCAA TTGCTGAATT AGTAGATCAA GCAACGCAAT TGAATCAATC	29880
50	TATGGATCAA TTACAACAAG CAGTTAATGA ACATGCTAAC GTTGAGCAAA CTGTAGATTA	29940
	CACACAAGCA GATTAGATA AACAAAATGC TTATAAACAA GCTATTGCTG ATGCTGAAAA	30000

55

TGCAAAACAA GCATTAAATG GTGATGAACG TGTAGCACTT GCTAAACAA ATGGTAAACA 30120  
 TGACATCGAC CAATTGAATG CATTAAACAA TGCTCAACAA GATGGATTTA AAGGTCGCAT 30180  
 5 CGATCAATCA AACGATTTAA ATCAAATCCA ACAAATTGTA GATGAGGCTA AGGCACTTAA 30240  
 TCGTGCAATG GATCAATTGT CACAAGAAAT CACTGACAAT GAAGGACGCA CGAAAGGTAG 30300  
 CACGAACTAT GTCAATGCAG ATACACAAGT CAAACAAGTA TATGATGAAA CGGTTGATAA 30360  
 10 AGCGAAACAA GCACTTGATA AATCGACTGG TCAAACTTA ACTGCAAAAC AAGTTATCAA 30420  
 ATTAAATGAT GCAGTCACTG CAGCTAAGAA AGCATTAAAT GGTGAAGAAA GACTTAATAA 30480  
 TCGTAAAGCT GAAGCATTAC AAAGATTGGA TCAATTAACA CATCTAAACA ATGCTCAAAG 30540  
 15 ACAATTAGCA ATCCAACAAA TTAATAATGC TGAAACGCTA AATAAAGCAT CTCGAGCAAT 30600  
 TAATAGAGCA ACTAAATTAG ATAATGCAAT GGGTTCAGTA CAACAATATA TTGACGAACA 30660  
 GCACCTTGGT GTTATCAGCA GCACAAATTA CATCAATGCA GATGACAATT TGAAAGCAAA 30720  
 20 TTATGATAAT GCAATTGCGA ATGCAGCACA TGAGTTAGAT AAAGTGCAAG GTAATGCAAT 30780  
 TGCaaaAGCT GAAGCAGAGC AATTGAAACA AAATATTATC GATGCTCAA ATGCATTAAA 30840  
 25 TGGAGACCAA AACCTTGCAA ATGCCAAAGA TAAAGCAAAT GCGTTTGTTA ATTCGTTAAA 30900  
 TGGATTAAAT CAACAGCAAC AAGATCTTGC ACATAAAGCA ATTAACAATG CCGATACTGT 30960  
 ATCAGATGTA ACAGATATTG TTAATAATCA AATTGACTTA AATGATGCAA TGGAAACATT 31020  
 30 GAAACATTTA GTTGACAATG AAATTCCAAA TGCAGAGCAA ACTGTCAATT ACCAAAACGC 31080  
 TGACGATAAT GCTAAA 31096

## (2) INFORMATION FOR SEQ ID NO: 60:

35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2243 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 40 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

45 ATGACAGAAT GGGAGCGAGG ACTTAGAATG TTTCCTAAAT CAGGTTTATT AAATTTTGAG 60  
 TTAGCGATAG TAAATCGTTC ATTAAATGAT GATGAAAAAG CATTAAAATA TGTGCGTAAA 120  
 GCATTAAATG CAGACCCTAA AAATACAGAT TATATTA ACT TAGAAAAAGA GTTGACTAAA 180  
 50 TCAAATGAGT CGAAAAATAA ATAACCTTTA TGATGTACAA CAGTTATTGA AAAGTTACGG 240  
 ATTTCTAATA TATTTTAAAA ATCCAGAAGA TATGTACGAA ATGATTCAAC AGGAGATTTC 300

55

	TAATCAGAGA AGGAATGAAC AGAAATGACA AAAATTATTT TAGCAGCTGA TGTAGGCGGG	420
	ACGACTTGTA AATTAGGTAT TTTCACACCT GAATTAGAAC AATTACATAA ATGGTCTATT	480
5	CACACTGATA CATCTGATAG TACAGGATAT ACACTTTTGA AAGGAATTTA TGATTCTGTT	540
	GTTGAAAAAG TAAATGAAAA TAATTATAAT TTTTCAAATG TACTTGGCGT AGGTATTGGT	600
	GTACCAGGTC CTGTTGACTT TGAAAAAGGT ACAGTAAATG GAGCAGTAAA CTTATATTGG	660
10	CCAGAAAAAG TTAATGTACG TGAGATTTTT GAACAATTCG TTGATTGTCC AGTGTATGTA	720
	GATAATGATG CTAACATAGC TGCTTTAGGG GaGAAACACA AAGGTGCTGG TGAAGGTGCC	780
	GATGATGTTG TTGCCATCAC ACTTGGTACA GGTCTAGGTG GAGGAATTAT TTCCAAATGG	840
	TGAAATCGTA CATGGTCATA ATGGCTCtGG CGCAGAAATA GGTCATTTTA GAgCAGACTT	900
	CgATCAACGA TTTaAATGTA ATTGTGGTCG TTCTGGATGT ATTGAAACAG TTGCTTCaGC	960
20	GACAGGCGTT GTTAACTTAG TTAACCTTcA CTATCCGAAG TTGACGTTTA GATCTTCTAT	1020
	ATTAGAATTG ATTAAAGAAA ATAAGGTtAC aGCAAAAGCT GTTTTTGATG CGGCAAAAGC	1080
	TGGTGACCAA TTCTGTATTT TCATTACTGA AAAGGTTGCA AACTATATTG GATATTTATG	1140
25	TAGTATTATT AGTGTTACAA GTAATCCGAA ATATATCGTT CTAGGTGGAG GAATGTCTAC	1200
	TGCAGGACCT ATTTTAATTG AAAATATTAA AACAGAATAT CATAATTTAA CATTTGCACC	1260
	TGCTCAATTT GAAACTGAAA TTGTACAAGC GAAATTAGGT AATGATGCAG GTATTACAGG	1320
30	AGCAGCAGGA TTAATCAAGA CCTATGTATT AGATAAAGAG GGGGTAAAAT AATGGCTATT	1380
	GTTGATGTGG TTGTTATTCC AGTTGGAACG GAAGGTCCGA GTGTTAGTAA ATATATTGCA	1440
	GATATTCAGA AAAAAGCTTCA AGAATATAAA GCAATGGGTA AAATTGATTT TCAATTAACA	1500
35	CCAATGAATA CTCTAATTGA AGGTGAATTA AGCGATGTAT TAGAAGTTGT GCAAGTGATA	1560
	CATGAATTAC CTTTGTATAA AGGTTTAAAGT AGAGTTTGTA CAAATATCCG TATTGATGAC	1620
	CGACGAGACA AATCTAGAAA AATGAATGAT AAACAAACAT CAGTACAAAA ACATTTAGAA	1680
40	AATAGTGGTG AAAACCTATG AGGATTTCAA GCTTAACTTT AGGCTTAGTT GATACTAATA	1740
	CGTATTTTCA CGAAAAATGAC AAAGCTGTTA TTCTGATTGA CCCTTCAGGT GAAAGTGAAA	1800
45	AAATTATTAA AAAATTAAAC CAAATAAATA AACCGTTAAA AGCTATTTTA TTAACACATG	1860
	CACACTTTGA TCATATCGGA GCAGTCGATG ATATAGTTGA TCGATTGAT GTCCCGGTTT	1920
	ATATGCATGA AGCAGAGTTT GATTTTCTAA AAGATCCCGT TAAAAATGGG GCAGATAAAT	1980
50	TTAAGCAATA TGGATTACCA ATTATTACAA GTAAGGTAAC TCCTGAAAAG TTAAmCGAAG	2040
	GTAGCACAGA AATAGAAGGA TTTAAGTTnT nAyrTGTaCA CACACCTGGA CATTCAACAG	2100
55		

GAATCGGACG TACAGATTTA TATAAAGGTG ATTATGAAAC GCTAGTTGAT TCTATTCAAG 2220  
 ATAAAAATATT TGAATTAGAA GGC 2243

5 (2) INFORMATION FOR SEQ ID NO: 61:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 8009 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

TTGGnATCAT tyAcgGTAAA AAGAATAAaG CAAGATTtAT TTCATTAGTA CTAATTTGTG 60  
 CAATGTTTGC AATTtGTTGG GTTGCATATA TTCAATGGGA GTCTACAATC GCTTCATTTA 120  
 CACAATCTAT TAATATTTCa ATGGCACAAT ATAGTGTtTT ATGGACAATT AACGGAATAA 180  
 TGATTTTtAGT AGCACAACCA TTAATTAAAC CGATTCTCTA TCTGTtAAAA GGAAACTTAA 240  
 AGAAGCAAAT GTTTGTcGGC ATCATCATTT TTATGTtGTC GTTCTTTGTC ACGAGTTTTG 300  
 CCGAAaACTT TACAATATTT GTTGTcGGTA TGATTATTTT AACTTTtGGA GAAATGTTTG 360  
 TATGGCCAGC AGTtCCAAct ATAGCCAATC AGTTAGCGCC AGATGGTAAG CAAGGACAGT 420  
 ACCAAGGTtT TGTGAATTCA GCTGCTACAG TAGGAAaAGC ATTTGGTCCA TTTCTTGGTG 480  
 GTGTATTAGT TGATCGGTTT AATATGCGCA TGATGTtTAT CGGTATGATG CTACTACTTG 540  
 TATTTGCATT AATATTATTA ATGGTTTtCA AGGAGAATAA TACGCAACCT AAAAAAATAG 600  
 ATGCATAATG AGTAAATAGA ATTAACGTTA TAGACTTGAA ATAAATGTcG TTATAACATA 660  
 ATATTAATTT GTATAATTtA ATTTcGTTTG GAGCTTTtCT ACAGAAaAGCT AGTGATGCTG 720  
 AGAGCTAGTG TTAAGGACTA AATGTAAATC GTATTAAATT TAAATTGAAT GAATGACATC 780  
 TCTTACTATT AAAATGAGTG CACAATTTTt GTGAAATAGG GTGGTAACGC GGCAAATGTC 840  
 GTCCCTATGT AAATAGAATA GTTAGAGGTG TCTTTTtTAT TGAATAGGAG GAAATGTGTT 900  
 GAATTACAAC CACAATCAAA TTGAAaAGAA ATGGcAAGAC TATTGGGACG AAAATAAAAC 960  
 ATTTAAaACA AATGATAACT TAGGTCAAAA GAAATTTTAT GCTTTAGACA TGTTTCCATA 1020  
 TCCATCAGGT GCTGGTTTtAC ATGTTGGACA TCCTGAGGGc TATACAGCAA CAGATATCAT 1080  
 TTCAAGATAT AAAAGAATGC AAGGATATAA TGTATTACAT CCGATGGGGT GGGATGCATT 1140  
 CGGATTACCA GCAGAGCAAT ATGCTTTAGA CACTGGCAAC GACCCACGTG AATTTACAAA 1200  
 GAAAAATATC CAAACTTTTA AACGACAAAT TAAAGAATTA GGGTTCAGTT ATGATTGGGA 1260

55

	GTTATATAAC AAAGGTTTAG CATACGTTGA TGAAGTTGCA GTTAACTGGT GTCCAGCATT	1380
	AGGCACTGTT TTATCTAACG AAGAAGTGAT TGATGGTGTC TCTGAACGTG GTGGACATCC	1440
5	AGTTTATCGT AAGCCGATGA AACAAATGGGT ACTTAAATC ACAGAATATG CAGATCAATT	1500
	ATTAGCAGAT TTAGATGATT TAGATTGGCC TGAGTCTTTA AAAGATATGC AGCGCAATTG	1560
	GATTGGACGT TCTGAAGGGG CCAAAGTTTC ATTTGATGTA GATAATACGG AAGGAAAAGT	1620
10	AGAAGTATTT ACGACTAGAC CAGATACAAT CTATGGTGCA TCATTCTTAG TCTTAAGTCC	1680
	TGAACATGCA TTAGTTAATT CAATTACAAC AGATGAATAT AAAGAAAAAG TAAAAGCTTA	1740
15	TCAAACAGAA GCTTCTAAAA AGTCAGATTT AGAACGTACA GATTTAGCAA AAGATAAATC	1800
	AGGTGTATTT ACTGGTGCAT ATGCAACTAA TCCTTTATCT GGTGAAAAAG TACAAATTTG	1860
	GATTGCTGAT TATGTATTAT CAACATATGG TACTGGAGCA ATTATGGCAG TACCAGCGCA	1920
20	TGATGACAGA GATTATGAAT TTGCTAAAAA GTTTGATTTG CCAATCATTG AAGTCATCGA	1980
	AGGTGGAAAT GTTGAAGAAG CAGCATACAC TGGTGAAGGT AAACATATTA ATTCTGGTGA	2040
	ACTTGATGGT TTAGAAAATG AAGCGGCAAT TACTAAAGCT ATTCAATTAT TAGAGCAAAA	2100
25	AGGTGCTGGC GAAAAGAAAG TTAATTACAA ATTAAGAGAT TGGTTATTCA GTCGTCAGCG	2160
	TTATTGGGGC GAACCAATTC CTGTCATTCA TTGGGAAGAT GGAACAATGA CAACTGTTCC	2220
	TGAAGAAGAG CTACCATTTG TGTTACCTGA AACAGATGAA ATCAAGCCAT CAGGGACTGG	2280
30	TGAGTCTCCA CTAGCTAATA TTGATTCATT TGTAATGTT GTAGATGAAA AAACAGGTAT	2340
	GAAAGGACGT CGTGAAACAA ATACAATGCC ACAATGGGCA GGTAGTTGTT GGTATTATTT	2400
	ACGTTACATC GATCCTAAAA ATGAAAATAT GTTAGCAGAT CCTGAAAAAT TAAAACATTG	2460
35	GTTACCTGTT GATTTATATA TCGGTGGAGT AGAACATGCG GTTCTTCACT TATTATATGC	2520
	AAGATTTTGG CATAAAGTCC TTTATGATTT GGCTATCGTA CCTACTAAAG AACCTTTCCA	2580
40	AAAATTATTT AACCAAGGTA TGATTTTAGG AGAAGGTAAT GAGAAGATGA GTAAATCTAA	2640
	AGGAAATGTA ATCAATCCTG ATGATATAGT ACAGTCTCAT GGTGCAGATA CTTTGCCTCT	2700
	TTACGAAATG TTTATGGGAC CTTTAGATGC TGCAATTGCA TGGAGTGAAA AAGGATTAGA	2760
45	TGGGTCTCGT CGATTCTTAG ATCGCGTATG GCGTTTAATG GTAAATGAAG ATGGGACATT	2820
	GAGTTCAAAA ATTGTAAC TAATAAATA ATCTTTAGAT AAAGTTTATA ACCAACTGT	2880
	TAAAAAGGTA ACAGAAGACT TTGAAACATT AGGATTTAAT ACTGCTATTA GTCAATTAAT	2940
50	GGTATTTATT AATGAGTGTT ATAAAGTTGA TGAAGTTTAT AAACCTTACA TTGAAGGCTT	3000
	CGTTAAAAATG TTAGCACCTA TTGCACCACA TATCGGTGAA GAATTATGGT CAAAATTAGG	3060

55

	TGATGAAGTA GAAATCGTTG TTCAAGTGAA TGGTAAATTG AGAGCTAAAA TTAAAATTGC	3180
	TAAAGATACA TCAAAAAGAAG AAATGCAAGA AATTGCCTTA TCTAATGACA ATGTTAAAGC	3240
5	GAGTATTGAA GGTAAAGACA TCATGAAAGT CATCGCTGTT CCTCAAAAAT TAGTCAATAT	3300
	TGTAGCTAAA TAATGTTTTA AGGAGGACTT TGAAATGAAG TCAATTACTA CAGATGAATT	3360
	AAAAAATAAA CTTTGTAGAAT CTAAACCAGT TCAAATTGTT GATGTTGCGTA CTGATGAAGA	3420
10	AACAGCAATG GGATATATTC CTAATGCAAA GTTAATTCCA ATGGATACCA TTCCGGATAA	3480
	TTTAAATTCA TTTAATAAAA ATGAAATATA TTATATTGTA TGTGCTGGTG GAGTTCGAAG	3540
	CGCTAAAGTT GTAGAATATT TAGAGGCAAA TGGCATTGAT GCCGTAAATG TCGAAGGCGG	3600
15	CATGCACGCA TGGGGCGATG AAGGTTTGGG AATAAAAAGT AITTTAAAGTA GTGACATAAT	3660
	TTAAAATAAT ATTACATTTG TAATGACACC AAGTAACGTT TCGGTTGCTT GGTGTTTTTT	3720
	GGTATGAATT ACTTCTGTT ACAAACAAT CTAAAGCGTT CTTGTTATGT TTTATTAAGA	3780
20	TTTTAATTAC AAAACGGAAA CTAAATTGTA ATAAAATAAA ACTTTATTTT ATAAAATGAT	3840
	GATGATAAAA TTGAGTGAAC TTAAAATATT GTACAAAATA ATATAGCTAT AAATATAATA	3900
25	TAGCTATAAA TATAATATGA GGGAGCGTAT ATTTTGTAGCA TAATTCTTAA CAACACAGCA	3960
	GAGAACAGAC AACCAGGAGG AAAATGAAAT GAATTTGTTA AAGAAAAATA AATATAGTAT	4020
	TAGGAAGTAT AAAGTAGGCA TATTCTCTAC TTTAATCGGA ACAGTTTTAT TACTTTCAAA	4080
30	CCCAAATGGT GCACAAGCCT TAACTACGGA TAATAATGTA CAAAGCGATA CTAATCAAGC	4140
	AACACCTGTA AATTCACAAG ATAAAGATGT TGCTAATAAT AGAGGTTTAG CAAATAGTGC	4200
	GCAGAATACA CCTAATCAAT CTGCAACAAC CAATCAAGCA ACGAATCAAG CATTGGTTAA	4260
35	TCATAATAAT GGTAGTATAG TAAATCAAGC TACGCCAACA TCAGTGCAAT CAAGTACGCC	4320
	TTCAACACAA AACAATAATC ATACAGATGG CAATACAACA GCAACTGAGA CAGTGTCAAA	4380
	CGCTAATAAT AATGATGTAG TGTGCAATAA TACCGCATT AATGTACCAA CTAAAACAAA	4440
40	TGAAAATGGT TCAGGAGGAC ATCTAACTTT AAAGGAAATT CAAGAAGATG TTCGTCATTC	4500
	TTCAAATAAA CCAGAGCTAG TTGCAATTGC TGAACCAGCA TCTAATAGAC CGAAAAAGAG	4560
45	AAGTAGACGT GCGGCACCGG CAGATCCTAA TGCAACTCCA GCAGATCCAG CGGCTGCAGC	4620
	GGTAGGAAAC GGTGGTGCAC CAGTTGCAAT TACAGCGCCA TATACGCCAA CAACTGATCC	4680
	TAATGCCAAT AATGCAGGAC AAAATGCACC TAACGAAGTG CTGTCATTTG ATGACAATGG	4740
50	TATTAGACCA AGTACCAACC GTTCTGTGCC AACAGTAAAC GTTGTTAATA ACTTGCCGGG	4800
	CTTCACACTA ATCAATGGTG GCAAAGTAGG GGTGTTTAGT CATGCAATGG TAAGAACGAG	4860

55

	TCGTATACAT	GGAAGTATA	CGAATGACCA	TGGCGATTTT	AATGGTATCG	AGAAAAGCATT	4980
	AACAGTAAAT	CCGAATTCTG	AATTAATCTT	TGAATTTAAT	ACAATGACTA	CTAAAAACGG	5040
5	TCAAGGCGCA	ACAAATGTTA	TTATCAAAAA	TGCTGATACT	AATGATACGA	TTGCTGAAAA	5100
	GACTGTTGAA	GGCGGTCCAA	CTTTGCGTTT	ATTTAAAGTA	CCTGATAATG	TGAGAAATCT	5160
	CAAAATTCAA	TTTGTACCTA	AAAATGACGC	AATAACAGAT	GCGCGTGGCA	TTTATCAACT	5220
10	AAAAGATGGT	TACAAATACT	ATAGCTTTGT	TGACTCTATC	GGACTTCATT	CTGGGTGACA	5280
	TGTTTTTGT	GAAAGACGAA	CAATGGATCC	AACAGCAACA	AATAATAAAG	AGTTTACTGT	5340
	AACAACATCA	TTAAAGAATA	ATGGTAATTC	TGGTGCTTCT	CTAGATACAA	ATGACTTTGT	5400
15	ATATCAAGTT	CAATTACCTG	AAGGTGTTGA	ATATGTGAAC	AATTCATTGA	CTAAAGATTT	5460
	TCCAAGTAAC	AATTCAGGCG	TTGATGTTAA	TGATATGAAT	GTTACATATG	ATGCAGCAAA	5520
20	TCGTGTGATA	ACAATTAAAA	GTACTGGAGG	AGGTACAGCA	AACTCTCCGG	CACGACTTAT	5580
	GCCTGATAAA	ATACTCGATT	TAAGATATAA	ATTACGTGTA	AATAATGTGC	CGACACCAAG	5640
	AACAGTAACA	TTTAACGAGA	CATTAACGTA	TAAACATAT	ACACAAGATT	TCATTAAATC	5700
25	AGCTGCAGAA	AGTCATACTG	TAAGTACAAA	TCCATATACT	ATCGATATCA	TCATGAATAA	5760
	AGATGCATTA	CAAGCCGAAG	TTGACAGACG	TATTCAACAA	GCTGATTATA	CATTTGCGTC	5820
	ATTAGATATC	TTTAATGGTC	TGAAACGACG	CGCACAAACG	ATTTTAGATG	AAAATCGTAA	5880
30	CAATGTACCA	TTAAATAAAA	GAGTTTCTCA	AGCATATATT	GATTCATTAA	CTAATCAAAT	5940
	GCAACATACG	TTAATTGAA	GTGTTGATGC	TGAAAATGCA	GTTAATAAAA	AAGTTGACCA	6000
	AATGGAAGAT	TTAGTTAATC	AAAATGATGA	ATTGACAGAT	GAAGAAAAAC	AAGCAGCAAT	6060
35	ACAAGTTATC	GAGGAACATA	AAAATGAAAT	AATTGGTAAT	ATTGGTGACC	AAACGACTGA	6120
	TGATGGCGTT	ACTAGAATCA	AAGATCAAGG	TATACAGACC	TTAAGTGGGG	ATACTGCAAC	6180
	ACCGGTTGTT	AAACCAAATG	CTAAAAAAGC	AATACGTGAT	AAAGCAACGA	AACAAAGGGA	6240
40	AATTATCAAT	GCAACACCAG	ATGCTACTGA	AGACGAGATT	CAAGATGCAC	TAAATCAATT	6300
	AGCTACGGAT	GAAACAGATG	CTATTGATAA	TGTTACGAAT	GCTACTACAA	ATGCTGACGT	6360
45	TGAAACAGCT	AAAAATAATG	GCATCAATAC	TATTGGAGCA	GTTGTTCTC	AAGTAACTCA	6420
	TAAAAAGCT	GCAAGAGATG	CAATTAACCA	AGCAACAGCA	ACGAAAAGAC	AACAAATAAA	6480
	TAGTAATAGA	GAAGCAACTC	AGGAAGAGAA	AAATGCAGCA	TTGAACGAAT	TAATCAAGC	6540
50	AACCAACCAT	GCTTTAGAAC	AAATCAATCA	AGCAACAACA	AATGCTAATG	TTGATAACGC	6600
	CAAAGGAGAT	GGTCTAAATG	CCATTAATCC	AATTGCTCCT	GTAAGTGTG	TTAAGCAAGC	6660

55



TGATGCGACT CAAGAAGAAA GACAAGCAGC AATTGACAAA GTGAATGCTG CTGTAAGTGC 6780  
 AGCAAACACA AACATTTTAA ACGCTAATAC CAATGCTGAT GTTGAACAAG TAAAGACAAA 6840  
 5 TGCGATTCAA GGAATACAAG CAATTACACC AGCTACAAAA GTAAAAACAG ATGCAAAAAA 6900  
 TGCCATCGAT AAAAGTGCGG AAACGCAACA TAATACGATA TTTAATAATA ATGATGCGAC 6960  
 GCTCGAAGAA CAACAAGCAG CACAACAATT ACTTGATCAA GCTGTAGCCA CAGCGAAGCA 7020  
 10 AAATATTAAT GCAGCAGATA CGAATCAAGA AGTTGCACAA GCAAAAGATC AGGGCACACA 7080  
 AAATATAGTA GTGATTCAAC CGGCAACACA AGTTAAAACG GATACTCGCA ATGTTGTAAA 7140  
 TGATAAAGCG CGAGAGGCGA TAACAAATAT CAATGCTACA ACTGGCGCGA CTCGAGAAGA 7200  
 15 GAAACAAGAA GCGATAAATC GTGTCAATAC ACTTAAAAAT AGAGCATTAA CTGATATTGG 7260  
 TGTGACGTCT ACTACTGCGA TGGTCAATAG TATTAGAGAC GATGCAGTCA ATCAAATCGG 7320  
 CGCAGTTCAA CCGCATGTAA CGAAGAAACA AACTGCTACA GGTGTATTAA ATGATTTAGC 7380  
 20 AACTGCTAAA AAGCAAGAAA TTAATCAAAA CACAAATGCA ACAACTGAAG AAAAGCAAGT 7440  
 GGCTTTAAAT CAAGTGGATC AAGAGTTAGC AACGGCAATT AATmATATAA ATCAAGCTGA 7500  
 25 TACAAATGCG GAAGTAGATC AAGCGCAACA ATTAGGTACA AAAGCAATTA ATGCGATTCA 7560  
 GCCAAATATT GTTAAAAAAC CTGCAGCATT AGCACAAATC AATCAGCATT ATAATGCTAA 7620  
 ATTAGCTGAA ATCAATGCTA CACCAGATGC AACGAATGAT GAGAAAAATG CTGCGATCAA 7680  
 30 TACTTTAAAT CAAGACAGAC AACAAAGCTAT TGAAAGTATT AAACAAGCTA ACACAAATGC 7740  
 AGAAGTAGAC CAAGCTGCGA CAGTAGCAGA GAATAATATC GATGCTGTTC AAGTTGATGT 7800  
 AGTAAAAAAA CAAGCAGCGC GAGATAAAAT CACTGCTGAA GTGGcGAacG TATTGaAGCG 7860  
 35 GTTAAACAAA CACCTAATGC AACTGACGAA GAAAAGCAGG CTGCTGTTAA TCAAATCCAA 7920  
 TCAACTTTAA AGATTCAAGC AATTTAATCC AAATTTAATC CAAAACCCAA ACAAATGGAT 7980  
 TCAGGGTAGG ACACCACTTA CAAATCCAA 8009

(2) INFORMATION FOR SEQ ID NO: 62:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10953 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

ACCCACCCCh TGGGGATAnT TTACCTGGTG GGGCCTTCGA TTGCCTTTAG GTGAAACCaG 60

	AGATGAATGC TAACCATATT CATTCTGCTA AAGATGGTCG TGTTACTGCG ACAGCTGAAA	180
	TTATTCATCG AGGTAAGTCG ACACATGTAT GGGATATAAA AATTAAGAAT GACAAAGAAC	240
5	AATTAATTAC AGTTATGCGT GGTACAGTTG CTATTAAACC TTTAAAATAA AAGAACTGCT	300
	AGCTGAAATG TTATGAGATA TTCATAACTA CGGCTAGCAG TTTTTTTATG CGCTATATTG	360
	TTGTAGTTTT AGAAATGCTT GTTCAATGCG TTCGGCAGCT TTACGGCCAC CCATAACATT	420
10	TCTACCAAAT GGTCCTAATT CTAAGTCTGC AAAGCATCCT GCGACAAATA GATTTGGTAT	480
	CCATTCTAAT TTTTCGGAAA TAACAGGGTA ATTACATTCG TTGATAGGTG CATCATAATT	540
	TTGTATTAAAT TGCTTAATAA GTGGTTGTGA CATAAAATCT TGTTCAAAAC CAGTTGCAAC	600
15	CATAATCTGT TGATATGGAA CAGAATCATT TTCAGTGTTA ATTACACCAC CACTAATTTG	660
	AGTGATAGGT GTTTTATGCa CATTIATACG ACCATTTTTTA ATATGTTTTT TAAGGCGTAA	720
20	GTACAGTTCG TGAGGCATTG ATCCTTTATG ACGTTCGCGT TGTACAATGG CATTTCTTTC	780
	AGGCATGCTT TTAGTACTTA AAAATGAAGA CATATTTTTTC GGACCTAACC AACCAGGATC	840
	AGCATCAAAG TCATGTATTT CAATATCTTT ATTTAGCCAT AAATGAATCT TTTTATCGTT	900
25	ATCATGATTT AACAAITTTAA GTGCAAGATG TGCAGCAGTa ATGCCGCTAC CAACGATATG	960
	ATCGGTCTTA TCATATACTA CTTGATCAAG TTCITTTCTCG AAGATATGAT TTACATTCTG	1020
	TTTGTCTTTT AAAATGTCAG GCATAAACGG AATATTTGTA CTGCCTATTG CAATAACGAC	1080
30	GCAATCTGTA GTGATAATTT GTCCATCTTC TAACTGATA TGCCATTTGT CTTCTTGTTT	1140
	ATCTAAAGTT TGAACATAAC CTGAAACCAA GCAATCCTCT AATTGATAAT GTTTAGAAGC	1200
	ATGTGCAATA TGATCCATAA ACATTGTCAA TTCAGGTCGT TGATAAGGAC CATAAAAAGC	1260
35	ATTTGTATAT TGGTGCTGTT TAGCGAATTG TTTTAGATGG AACGGTTGTG GATGTACGTG	1320
	ATGTACAATC GGTGATCTTA AATAAGGCAT TTCTATTCGA TTTGTATATG AGTTAAACCT	1380
	TTGGCAAAAA GTTTCGTGTG GGTCAATGAT TGTTAATCGG TCTGTTGTTA ATCCGCTTGA	1440
40	TAATAGTTTT TGTGCGATTG CAGTTCCCTG TATGCCACCG CCGATAATTG TCCAATGCAT	1500
	AATAAACCT CTCTCTTTTT AAAACGTAAT AGTTACGATT TATAATTATT ATTATCATAA	1560
45	TACATAACGA CATGAAAGGC AATTAAATTA AAGAGATATA TGTAGATAGG GCGAATCTGT	1620
	AGTCAAAGAA AAAATCATTG AAAAAGAGGT AACATGTCA AAAGAWAACA GCAGTAAAT	1680
	CATTCCTAAT TTGGAATCAT CTTACTGCTG TTTGTTGTTG ATTTATATTG ATGATTTTGT	1740
50	TATATAATCT ACAATTTTGT GTCTTTTAAG TCTTCCGAAA TTTCATCGAC TTTAGTCTTT	1800
	TTAGTATAAG GCGTTTTAAT ATTATATGCT GCTTTCATAA TCATATGACT TGAAAGAGGA	1860

55

	GCAATAAAAT ATAAAAACGT ACCAAATAGT AATGACATTG CACCTAATGT TGATGCTTTT	1980
	CCGGCAGCAT GTGCACGTGA ATATACATCT TCAAGTCTCA ATAATCCTAT AGCTGCTAGG	2040
5	GCGCTAATTA AAGCACCGAT GATAACAAAG ATAAGTGCAA GACTAATCAG TATGATTTTG	2100
	ATCATGTTCA ATCACCCTTAC CTTTGTCCAT AAATTTAGAG AATACTGCAG TACCTAAAAA	2160
	AGCTAATATA CCAATCATCA TAATAACGAC AATCATGTAT TTAATATTTA ATAAAATACT	2220
10	GAATAATGCT ATAACTGCCA TTAATTGAAG ACCAATCGCA TCTAATGCGA CAACACGATC	2280
	GGCAAGTGAT GGGCCTAGCA CAACGCGAAT GAGCATAGCT AACATAGAAA TGACAACATAT	2340
15	GATTAATGCA ATAACGATAA TAACATTATG ATTCATTATA TTTCGCCAC CTCTCTTACA	2400
	ATTTTCTCTA ATGATGTTTT AATACTTTCT ACTTCTTGCT CTTTAGTTGA AAAATCTATG	2460
	GCATGAATAT AAATTTTTGT ACGATCGTCA CTTACACCAA GCACTACAGT ACCAGGTGTT	2520
20	AATGTAATTA AATTAGACAG CAAGACAATT TGCCAATCTT TTTTAAATC TGTGTGATAA	2580
	ACAAAGAATC CTGGTTTATT TTTAATCGAA GGTTTAATAA TAATTTTCAA AACATCAAAA	2640
	TTAGCTTTAA TCAGTTCGAT TAAGAAAATA ATAACATAAT TAATAATACG ATATAGCGTG	2700
25	ATGACATAAA ATCTACCTGG TAACACTCTG TGTAAGAGGT AAACAAGAAC TAGGCCAAAG	2760
	ATGAAACCTA ACACAAAGTT ATTTGTTGTG TAACTATTTG TCACAAACAA CCAAAACACT	2820
	GCGATAATAA AGTTTAATAC TAATTGTACA GCCATGTTAT TTACCTCCTA ATACAGCTTT	2880
30	AACGTAGGTT GATGGATTGT AGAATGTTT TGCACCAGCT TTTACCATTG GATATAAGTA	2940
	ATCTGCTGAC AATCCATATA AAACAGTTAT CACAACTGCA ACGATTGCAA TCGTAGTTAA	3000
	ATATTTGACG TCGACTTTGT TATTAAGATC ATATCCTTTT GGTTGACCGA AAAAGCCTTG	3060
35	TAGGAATATG CGAATGACAG AATATAATAC GACTAAACTT GATAATAAGA CGATGACACC	3120
	ACTTAAATAA AATCCTCTTT CAAATGTTGA TTGGACAATA AAAAATTTTC CATAAAAGCC	3180
	ACTGAGTGGG GGAATGCCAG CTAAACTTAA TGCTGCGATA AAGAATGACC AACCAAGTAC	3240
40	AGGATATCGT TTAATTAAGC CACCAAATTG TCTTAAATCA GCAGTGCCTG TAATTTTAAT	3300
	CATAATTCCG ATAAGCAAGA ATAATGCAAG TTTTACTAAC ATGTCGTGCA ATGTATAGTA	3360
45	AATAGCCCCA ATCATACCTG ACTCTGTCAT CATTGCAACG CCGACTAAGA TCACACCTAC	3420
	AGCAATCATG ACATTGTATA GGATGATTTT TTTAATGTTG GCATATGCAA CAGCACCGAC	3480
	ACAACCAAAG ATGATCGTTA ATAGTGCTAA GAATAAAATG ACATAATGTG AAAAGCTTAC	3540
50	ATTATCACTA AAGAATAGGC TCAATGTTCT AGCGATTGCA TAAACACCAA CTTTTGTTAA	3600
	CAAAGCACCA AAGAATGCAA TGATTGGAAT TGGTGGCAT AGTATGCACT AGGTAACCAA	3660

55

	ATATTGACTA AGCCACTGTC ATGCGCTGAA AGGTTAGCTA ATTTATTGCT TATATCTGCT	3780
	AGATTCAATG TTCCTACTAC TGAATATAAA ATCGCTACAC CCATTACGAA GAAGGATGAC	3840
5	GATACAACGT TAACAAGAAC ATATTTTATT GTTTCTTGTA GTTGAATTTT TGTAGAACCA	3900
	ATTACTAATA AGAAATAAGA TGACATTAAA AATACTTCGA AAAATACGAA TAGGTTGAAA	3960
	ATGTCACCAG TTGTGAATGC ACCAATGATA CCTATTAAAC TAAATAGTAC TGAAAAATAA	4020
10	TAATAATATC TTTCACGTTT AATACCAATT GTTTGGTATG AATATAAAAT CACAATAGCT	4080
	GTAATAATAA TACTAGTAAT TATTAGTAGG GCACTGAATA TGTCTAATAC AAAGACAATA	4140
	CTGTATGGTG CTTTCCATGA ACCTAGCTCT ACGCGTATTG GTCCATGTTT AACACATTT	4200
15	GCTAAATTGA TAATTGCCGC GACCAAGGTT AATAATGTAC CGCCTAGTGC GACATAACGC	4260
	TTTATAATAG GACGCTTTCC AATAAAGACA AGTAATATGG CTGTAATTAC TGGAATAACT	4320
	AGCGTTAACA CAAGCATATT ACTTTCAATC ATCTTCTGGA ACTCCTTTCA TACTCTCAAC	4380
20	GTATCTGTG CCTAATTCTT TATATGTTCT AAATGCTAAT ACTAAGAAAA AGGCTGTGT	4440
	CGCAAGCGCA TAACGATTGC TGTTAAAATA AGTGCTTGCG GGaTAGGaTC AACATAGCTT	4500
25	TTTACGTTCC CTTCATAAAT TGGAACAGTA CCATGTTTAA GTCCGCCCAT AGTTATTAAA	4560
	AATAAAATTG CTGCATGTGT TAATAGTGTA GTTCCCATAA CAATTCGTAT CAGACTTTTA	4620
	GACAAAACGA GATAGACACT AATGCTGTG AGAATACCAC TAACAAAAAT CATAATAATT	4680
30	TCCACTATTC GTTCTCTCCA ATCGAAATAA TAATTGTCAT GACAGTACCA ACTACTGCAC	4740
	ATAAACACC GAAATCAAAG AATACTGCTG TTGTCATATG AACAGGTTCT AATATAAATA	4800
	ACGGTATATC AAATGTGACA TGCGTAAAGA AATTTTTGCC TAAAAACCA CTTGCGATAG	4860
35	GCGTCGCAAT ACAAAAAACT AATCCGATAC CTATCAAGAT TTTAAATCT AATGGGAAAA	4920
	TTTTACGCAT TGTTTCTATA TCAAATGCAA TCGTAATGAT AACAAAGTGAA CTTGCGAATA	4980
	ATAATCCGCC GACGAAACCG CCACCAGGTG TATAATGTCC TGCTAAGAAA AGTGAAAAAC	5040
40	CAAAGACCAT TACCATGAAA AAGATAATAA CTGCAGCAAA TTGCAAAATT AGATCATTTT	5100
	GTTGTCTATT CATGATTTTT CACCTCGTTA CCTTGCGTTT GACGCTTTTT ACGTAATTTA	5160
45	ATCATTGTAT ATACAGCTAA TCCTGCGATA CCAAGCACAG ATGACTCGAA TAAAGTATCC	5220
	ATACCACGGA AATCAACAAG TATGACGTTT ACCATGTTTT TACCGTGAGC tAAATCATAA	5280
	ACGTGCTCTT GATAAAACTT AGATATCGAT TCAAAATGTC TATTTCCGTA TGCAATTAAA	5340
50	CCGATAATAA TGACGGACAA ACCAACACCA CCAGCAATTA AAGCATTAGT AAGCTGGAAT	5400
	GAGCGCTTTT CATTATAACG ATTTAAATTT GGTAAGTGGT AGAAGCATAA TAAGAACAAT	5460

55

ATAAACAATA CAGACACAGC ATATCCAAC T GCACTTAACA TAATGATGCT AAATAATCTT 5580  
 GATTTAGCGA AAAGAATTAA AAAGGCAGCA CTTAATAATA AAATTACGAT ACAAACCTCG 5640  
 5 AAAATTCTAA TCGGACTAAC GTCTTTAAAA TTAATGTTGA AAGGTACTGA GAATATAGTG 5700  
 ACAAATGTTA ATAAAATTAA TGCACCAAAA ATGATAACTA AATTATTACG TGAATAATCG 5760  
 GTAACATAGC TATTCGTCAT CTTTTCAGAG TAGTTTGGAA TAACATTTGC ACTTCTGTTG 5820  
 10 TACCAATAAT TGAATGTTAG TTTACCAGGT TGTGCTTGCA ACAATTTTAC CCAATAACTA 5880  
 AATGTCACAA TTAGTAAGAT ACCTAAAATA TAAATCACTA ATGTTGATAA AAAGGCAGGC 5940  
 GTTAATCCAT GGAACATATG GAATTCAACA TCATCAATTA CCGTATGATT AATCGAAGag 6000  
 15 ThAGCTGGTT CAATAATCGA ATTAGTTAAA ATGCCAGGGA ATAAACCAAA TACAATTACT 6060  
 AATGTAGCTA AAATAGCTGG TGATAAAAGC ATTAATATTG ATACTTCGTG TGCTTTTTTA 6120  
 GGTAATTGTT CAGGTTTATA TTGTCCGAAA AATATATGCA TTATAAATTT AATTGAATAT 6180  
 20 ACAAATGTGA AGACACTGCC CACTATACCA ATGATTGGGA ATAGGTAGCC TAATGTATCA 6240  
 ACACTGAATA AATTTGCTTG GCTTGCTGTA AATGTTGTTT CTAAAAATGA TTCTTTTGAT 6300  
 25 AAGAAACCAT TGAACGGTGG TACACCAGCg CATACTTAAT GCTGTAATAA CAGTGATTGT 6360  
 AAATGAAATA GGCATAATTG TTAGTAAGCC ACCTAATTTT TTAACATCAC GTGTACCAGT 6420  
 AGAATGATCC ACTGCACCTG TAATCATAAA TAGGGCACCT TTAAATGTTG CATGGTTGAT 6480  
 30 TAAATGGAAT ATTGCAGCCG TAAATGCAGC AGCATATATT TTGCTATCAT CGCCTTGATA 6540  
 GTGATAACTA ATGGCACCGA TTCCAAGCAT CGCCATAATC ATACCTAATT GGGATACTGT 6600  
 TGAAAATGCC AGTATACCTT TCAAGTCTTG TTGTTTTGTT GCGTTTAGCG AAgCCCAGAA 6660  
 35 TAATGTAATT AAACCAACGA GTGTGACAGT CCATACCCAA CCTTGCGATG CTGCGAAGAT 6720  
 TGGTGTcATT CGAGCGATTA AATATAACCC TGCTTTAACC ATTGTTGCTG AATGAAGATA 6780  
 AGCACTGACT GGTGTAGGTG CTTCCATTGC ATCTGGTAGC CAAATATAAA ATGGAAACTG 6840  
 40 AGCAGATTTT GTAAAAGCAC CAATCATGAT TAAATCATC GCAAAAATGA AGAATGGGCT 6900  
 ATTTTGAATT TCAGAAGCAT GTTGAATCAT GTACTGAATG CTAAATGATT GTGTTGGTAT 6960  
 45 AGCGAGTAAG ATGATACCAC CTAATAATGA TAGACCACCA AATACTGTGA TTATGAGCGA 7020  
 TTTTGTAGCA CCATATATAG ATGCTTGTCG TTCGCGCCAG AATGAAATAA GTAAAAAACT 7080  
 AGAAAAAGAC GTTAGCTCCC AGAATAAATA TAGAATAATA ACATTATCTG AAAGTACGAC 7140  
 50 ACCTAACATT GCACCCATAA ATAGTAATAA ATAACAATAA AAATTCCTTA GTTGTCTGA 7200  
 CTTACTTAAG TAGCCGATTG AATATAATAC TACTAACTG CCGATTCTG AAATAAGCAA 7260

55

CCAATTTAAG GTTTTCATTA CAGTATTACC TGACATCGTC GTTTTAATTA ATGTAAGCAT 7380  
 ATAAATAAAT ATGACGATAG GGACAGGTAA TACGAACCAT CCTAAATGTA TACGTTTAAA 7440  
 5 AAATCTATAC AGGATAGGAA TAATGAGTGC GAATATTAAC GGTAATATCA CCGCAATATG 7500  
 TAACAAACTC ACTATGTTGT CCTCCTTTAA AAAATATTTA TGTTATTTCAT TATACATGAA 7560  
 TGATATAGTT CTGAAAAACG TACACACTCC TTGTTGTGCT TTATTTTCAG AaGTATTTAA 7620  
 10 ATAAGAAGAA ACACGTCATT TTTTATTTAA AATTTTCTTT GTATTGAAGT GAATAATCTT 7680  
 CTTTTAAGCG TGCTAAACTA GCTAAAGACA TTTCAGCATG TTTTGTGTTG TGAGCTTTAA 7740  
 GTTTAGTTTC TAAATCTGTA ATTGCTTGTT GAAGTGAATC TTCATAGCGC AATACATCAA 7800  
 15 CATTGAAGTC GCGTAATTGT GAACGTTTCG TATAGCGTTT TTCAAAATGG CTTAATGCTT 7860  
 TGCGGTCATG GAAAAATACA CCTTCAGTTT CAGTAGGGTT ATGTAAATCA CCTTGTTTCG 7920  
 20 GGTGTTTGAT AACTTGTTCA ACTTTAACAA GGACATCGTC TCCATTTTCT TCAACAATCG 7980  
 TGACACCATA GCTACCTGTT TTGTGTGAAA ATCGATATAG CTTCATGCTA TTTTCCTCCC 8040  
 TTAAAAGTAT GTTAATATAT ATGTATCATA ACATGAATGG AGAATATAAA TGGCTAACTA 8100  
 25 TCCACAGTTA AACAAAGAAG TACAACAAGG TGAAATCAAA GTGGTTATGC ACACAAATAA 8160  
 AGGTGACATG ACATTCAAAT TATTTCCAAA TATTGCACCA AAAACAGTTG AAAATTTTGT 8220  
 GACACATGCA AAAAATGGTT ATTATGAIGG AATCACATTC CACCGTGTCA TTAATGACTT 8280  
 30 CATGATTCAA GGTGGCGATC CAACAGCTAC TGGTATGGGT GGCGAAAGTA TTTATGGCGG 8340  
 TGCTTTTGAA GATGAATTTT CATTAAATGC ATTTAACTTA TATGGCGCAT TATCAATGGC 8400  
 TAACTCAGGA CCTAATACTA ATGGTTCACA ATTTTTCATT GTTCAAATGA AAGAAGTACC 8460  
 35 TCAAAATATG TTAAGTCAAC TTGCAGATGG TGGCTGGCCT CAACCAATCG TTGATGCATA 8520  
 TGCGGAAAAG GGTGGTACAC CATGGTTAGA TCAAAAACAT ACAGTATTCG GTCAAATCAT 8580  
 TGATGGTGAA aCTACATTAG AAGATATTGC AAATACAAA GTGGGACCAC AAGATAAACC 8640  
 40 ACTTCATGAT GTTGTAATTG AATCTATTGA TGTTGAAGAA TAATATCTAA ACATAATTAA 8700  
 CTACCAACAT TTAAACTCG GATAAAGCTA ATTTATGAAT GGATTAGTAT ATATTCCAAC 8760  
 45 gAAAAATAAAT AAATAATAT GATGAGCAAT CTCAATATAT TTATCaAGAA AGCACAGTTT 8820  
 TTAAATAGAT GTGTATTTTA AAGATAATAG TTGAGGTTGC TTTTATGTT TTTACAGAGA 8880  
 ATTGCTATTC AAATAGTAAA TAAATTGAAA ACAAAGTAGC TGGATATCAT ATTGATTTAG 8940  
 50 ATAGGAATTT GTTGCTAATT TTATTTGTAA ATCCAAGTTT GTAGAATTCT TATTCATTTA 9000  
 TAAATAATA TTCGTATGAT TTGATTTTTT AATTAGTCCA CCATTTGAT TTGTGCTATG 9060

55

## EP 0 786 519 A2

	AACATATCAA GGTGCGTGTA CTGGTATTCA ACCATACGGT GCGTTTGTG AGACCCCTAA	9180
	TCATACTGAA GGACTGATTC ATATATCAGA AATTATGGAT GACTACGTTC ATAATTTGAA	9240
5	GAAATTTCTA TCAGAAGGCC AAATTGTTAA AGCTAAAATT TTGTCTATAG ATGATGAAGG	9300
	AAAGCTTAAT CTATCATTA AGGATAATGA TTA CTTCAAA AATTATGAGC GTAAGAAGGA	9360
	AAAACAATCA GTATTAGATG AAATCAGAGA AACAGAAAAA TATGGGTTTC AAACACTTAA	9420
10	AGAACGCTTA CCAATCTGGA TAAAACAGTC AAAGCGAGCA ATTCGAAACG ACTAAAGGAA	9480
	CAGATAAATC GTACCGAAAA TCATACAAAG GGTCTGAAAT GAAAGTTTCT TAGACTATAA	9540
	AAGAGATTAG TATCTATTAA ATTTTATTAG ATACTAATCT CTTTTTGTCT ACGATAACGT	9600
15	AATATGaTTG ATTCTATTTA CACGTACAAA TGGTTTAAGG TGACATATCC ATTATCTTTG	9660
	TTAGATAGAA TCGTTGATTT GCaATATTGT ATGTGGATTT GTTTTTTTTA TTTATTTTAG	9720
	AAATGAGAAC TACAACTTAA AGTATTAAAC GAATTGCAAC TATATAAACA GATAATTGGA	9780
20	GAATGAAAAA ATTACATGTT ATAGTCAACT CAATAATTTT AAGGAGGAAT TAAGTAATGA	9840
	AAAGTAAATA CGAACCATTG TTTGATAAAG TAGAATTACC AAATGGAGTA GAGTTGAGAA	9900
25	ATCGATTTGT GTTAGCCCCCT TTAACACATA TTTCTTCAAA TGATGATGGT ACTATTTTCAG	9960
	ATGTAGAACT TCCTTATATT GAAAAGCGTT CACAAGATGT TGGTATTACA ATTAATGCTG	10020
	CGAGTAATGT GAGTGATGTC GGAAAAGCAT TTCCAGGACA GCCATCAATC GCGCATGACA	10080
30	GTAATATTGA AGGACTAAAA CGATTAGCTA CAGCAATGAA GAAAAACGGT GCCAAAGCAC	10140
	TCGTACAAAT ACATCATGGC GGTGCACAAG CATTGCCTGA ATTAACACCT GATGGAGACG	10200
	TCGTAGCACC AAGTCCAATT TCTTTAAAAA GTTTTGGTCA GAAACAAGAA CATAGTGCTA	10260
35	GAGAAATGAC GAATGAAGAG ATTGAACAAG CAATCAAGGA TTTTGGTGAA GCAACGCGAC	10320
	GTGCAATTGA AGCAGGGTTT GATGGTGTG AAATACATGG CGCGAATCAT TACTTAATTC	10380
	ATCAATTTGT ATCACCATAC TATAATAGAA GAAATGATGT ATGGGCAAAT CAATATAAAT	10440
40	TCCCGTGCGC TGTGATTGAA GAAGTACTTA AAGCGAAAGA AGCGTATGGC AATAAAGACT	10500
	TTATAGTTGG ATACAGATTA TCTCCAGAGG AAGCGGAGTC TCCAGGAATC ACAATGGAAA	10560
45	TTACAGAGGA ACTCGTTAAT AAAATTAGCC ATATGCCAAT CGACTATATT CATGTTTCAA	10620
	TGATGGATAC GCATGCAACG ACACGTGAAG GTAAATACGC TGGACAAGAA AGACTGCCTT	10680
	TAATTCACAA ATGGATAAAT GGTGCTATGC CACTTATCGG TATTGGTTCA ATTTTCACAG	10740
50	CTGACGAAGC TTTAGATGCA GTTGAAAATG TTGGTGTGA CTTAGTAGCC ATTGGTAGAG	10800
	AGCTIACTACT GGATTATCAA TTTGTTGAAA AAATTAAAGA TGGACGGGAA GATGAAATTA	10860

55

AATTTAATGA AGGGTTTTAT CCATTACCAC GTA

10953

(2) INFORMATION FOR SEQ ID NO: 63:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 8155 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

TTTGATAnAA AACTGAATnA ATTAAATGTA TCGATTCAAC CTAATGAAGT GAATTTACAA	60
GTAAAGTAG AGCCTTTTAG CAnAAAGGTT AAAGTAAATG TTAAACAGAA AGGTAGTTTA	120
GCAGATGATA AAGAGTTAAG TTCGATTGAT TTAGAAGATA AAGAAATTGA AATCTTCGGT	180
AGTCGAGATG ACTTACAAAA TATAAGCGAA GTTGATGCAG AAGTAGATTT AGATGGTATT	240
TCAGAATCAA CTGAAAAGAC TGTA AAAATC AATTTWCCAG AACATGTCAC TAAAGCACAA	300
CCAAGTGAAA CGmAGGCTTA TATAAATGTA AAATAAATAG CTAAATTAAA GGAGAGTAAA	360
CAATGGGAAA ATATTTTGGT ACAGACGGAg TAAGAGGTGT CGCAAACCAA GAACTAACAC	420
CTGAATTGGC ATTTAAATTA GGAAGATACG GTGGCTATGT TCTAGCaCAT AATAAAGGTG	480
AAAAACACCC ACGTGACTT GTAGGTCGCG ATACTAGAGT TTCAGGTGAA ATGTTAGAAT	540
CAGCATTAAAT AGCTGGTTTG ATTTCAATTG GTGCAGAAGT GATGCGATTA GGTATTATTT	600
CAACACCAGG TGTTGCATAT TTAACACGCG ATATGGGTGC AGAGTTAGGT GTAATGATTT	660
CAGCCTCTCA TAATCCAGTT GCAGATAATG GTATTAAATT CTTTGATCA GATGGTTTTA	720
AACTATCAGA TGAACAAGAA AATGAAATTG AAGCAATTATT GGATCAAGAA AACCCAGAAT	780
TACCAAGACC AGTTGGCAAT GATATTGTAC ATTATTCAGA TTACTTTGAA GGGGCACAAA	840
AATATTTGAG CTATTTAAAA TCAACAGTAG ATGTTAACTT TGAAGGTTTG AAAATTGCTT	900
TAGATGGTGC AAATGGTTCA ACATCATCAC TAGCGCCATT CTTATTTGGT GACTTAGAAG	960
CAGATACTGA AACAATTGGA TGTAGTCCTG ATGGATATAA TATCAATGAG AAATGTGGCT	1020
CTACACATCC TGAAAAATTA GCTGAAAAAG TAGTTGAAAC TGAAAGTGAT TTTGGGTTAG	1080
CATTTGACGG CGATGGAGAC AGAATCATAG CAGTAGATGA GAATGGTCAA ATCGTTGACG	1140
GTGACCAAAT TATGTTTATT ATTGGTCAAG AAATGCATAA AAATCAAGAA TTGAATAATG	1200
ACATGATTGT TTCTACTGTT ATGAGTAATT TAGGTTTTTA CAAAGCGCTT GAACAAGAAG	1260
GAATTAAATC TAATAAAACT AAAGTTGGCG ACAGATATGT AGTAGAAGAA ATGCGTCGCG	1320



	CTGGTGATGG TTTATTAAC T GGTATTCAAT TAGCTTCTGT AATAAAAAATG ACTGGTAAAT	1440
	CACTAAGTGA ATTAGCTGGA CAAATGAAAA AATATCCACA ATCATTAAAT AACGTACGCG	1500
5	TAACAGATAA ATATCGTGTT GAAGAAAATG TTGACGTTAA AGAAGTTATG ACTAAAGTAG	1560
	AAGTAGAAAT GAATGGAGAA GGTGCAATTT TAGTAAGACC TTCTGGAACA aACCATTAGT	1620
	TCGTGTCATG GTTGAAGCAG CAACTGATGA AGATGCTGAA aGATTTGCAC AACAAATAGC	1680
10	TGATGTGGTT CAAGATAAAA TGGGATTAGA TAAATAAATA CTGTATTACA AATGAGCCGA	1740
	TGCGTATGcA nTcgtTTTTT GTGTTTGTAG AAATAATTTA TAGTACAAAC GTAAAATGAT	1800
15	ATAAACAAAA TAAAAACAAA GTAATCAATA TGTAATATAA AATACACTGG TACTCAATAT	1860
	ATAATGATGA TAAATTAAT TTTAATTAGA TAGAGTTGCT TTGTGTTTTT AACGCAGATG	1920
	CTACTACTTA TCTTAACAGT TGATTAAGTG AAATCATTTA ACAGCGAGAA TAATCAACCA	1980
20	GGAGGATGAC TTAATGAATT TATTCAGACA ACAAAAATTT AGTATCAGAA AATTTAATGT	2040
	CGGTATTTTT TCAGCTTTAA TTGCCACTGT TACTTTTATA TCTACTAACC CGACAACAGC	2100
	GTCTGCAGCA GAGCAAAATC AGCCTGCACA AAATCAACCA GCACAACCAG CTGATGCCAA	2160
25	TACACAGCCT AACGCAAATG CTGGTGCTCA AGCTAATCCT ACAGCACAGC CAGCTGCACC	2220
	TGCCAACCAA GGACAACCAG CAGTACAACC AGCAAACCAA GGTGGACAGG CTAATCCAGC	2280
	AGGAGGAGCA GCACAACCAA ATACACAACC AGCTGGACAA GGTGATCAAG CTGATCCGAA	2340
30	TAACGCTGCA CAAGCACAACT CTGGAAATCA AGCAACACCG GCAAACCAAG CAGGTCAAGG	2400
	AAATAACCAA GCAACACCTA ATAATAATGC AACACCGGCA AATCAAACAC AGCCAGCGAA	2460
	TGCTCCAGCA GCAGCGCAAC CAGCAGCACC TGTAGCAGCA AACGCACAAA CTCAAGATCC	2520
35	AAATGCTAGC AATACTGGTG AAGGCAGTAT TAATACGACA TTAACATTTG ATGATCCTGC	2580
	CATATCAACA GATGAGAATA GACAGGATCC AACTGTAAC TTTACAGATA AAGTAAATGG	2640
40	TTATTCATTA ATTAACAACG GTAAGATTGG TTTCGTTAAC TCAGAATTAA GACGAAGCGA	2700
	TATGTTTGAT AAGAATAACC CTCAAAATA TCAAGCTAAA GGAAACGTGG CTGCATTAGG	2760
	TCGTGTGAAT GCAAATGATT CTACAGATCA TGCTAACTTT AACGGTATTT CAAAACTGT	2820
45	AAATGTAAAA CCAGATTCAG AATTAATTAT TAACTTTACT ACTATGCAAA CGAATAGTAA	2880
	GCAAGGTGCA ACAAATTTAG TTATTAAAGA TGCTAAGAAA AATACTGAAT TAGCAACTGT	2940
	AAATGTTGCT AAGACTGGTA CTGCACATTT ATTAAAGTA CCAACTGATG CTGATCGTTT	3000
50	AGATTTACAA TTTATTCCTG ACAATACAGC AGTTGCTGAT GCTTCAAGAA TTACAACAAA	3060
	TAAAGATGGT TATAAATACT ATTCATTCAT TGATAATGTA GGTCTATTCT CAGGATCACA	3120

55

	TAATACTGAA ATCGGTAACA ATGGTAATTT TGGTGCTTCA TTAAGAGCAG ATCAATTTAA	3240
	ATATGAAGTA ACATTACCAC AAGGTGTAAC TTACGTTAAT AATTCATTAA CTACAACATT	3300
5	CCCTAATGGT AATGAAGACA GTACAGTATT GAAAAATATG ACTGTTAATT ATGATCAAAA	3360
	TGCAAAATAAA GTTACATTTA CAAGCCAAGG TGTGACAACG GCACGTGGTA CACACACTAA	3420
	AGAAGTTTTA TTCCAGATA AATCTTTAAA ATTATCATAT AAAGTTAATG TTGCGAATAT	3480
10	CGATACACCT AAAAAATATTG ATTTTAATGA AAAATTAACA TATCGTACTG CTTCAGATGT	3540
	TGTAATTAAT AATGCGCAAC CAGAAGTaCA CTAAGTGCAG ATCCATTTTC AGTAGCGGTT	3600
	GAAATGAACA AAGATGCGTT GCAACAACAA GTAACTCAC AAGTTGATAA TAGTCATTAC	3660
15	ACAACAGCAT CAATTGCAGA ATACAATAAA CTAAACAAC AAGCAGATAC TATTTTAAAT	3720
	GAAGATGCGA ATCATGTTAA AACTGCAAAT CGTGCACTC AAGCGGATAT TGATGGTTTA	3780
	GTAACATAAT TACAAGCTGC ATTAATTGAT AATCAAGCAG CAATTGCTGA ATTAGATACT	3840
20	AAAGTCAAG AAAAGGTTAC AGCAGCACAA CAAAGTAAAA AAGTTACGCA AGATGAAGTT	3900
	GCAGCACTTG TAACTAAAA TAACAATGAT AAAATAATG CAATCGCAGA AATTAATAAA	3960
25	CAAACTACAG CACAAGGTGT CACAAGTAA AAAGATAATG GTATCGCAGT GTTAGAACAA	4020
	GATGTGATTA CACCAACAGT TAAACCTCAA GCGAAACAAG ATATTATCCA AGCAGTTACA	4080
	ACTCGTAAAC AACAAATTAA AAAGTCAAAT GCATCATTAC AAGATGAAAA AGATGTAGCA	4140
30	AATGATAAAA TTGGTAAAA TGAAACAAAG GCAATTAAAG ATATTGATGC AGCAACAACA	4200
	AATGCACAAG TAGAAGCCAT TAAACAAAA GCAATCAATG ATATTAATCA AACTACACCT	4260
	GCTACAACAG CTAAAGCAGC AGCTCTTGAA GAATTTGACG AAGTTGTTCA AGCACAAATT	4320
35	GATCAAGCAC CTTTAAATCC TGATACAACA AATGAAGAAG TAGCGGAAGC TATTGAACCT	4380
	ATTAATGCAG CTAAAGTTTC TGGTGTAAA GCAATTGAAG CGACAACGAC TGCACAAGAT	4440
	TTAGAAAGAG TTAACAAACGA AGAAATCTCA AAAATTGAAA ATATTACTGA CTCTACGCAA	4500
40	ACAAAAATGG ATGCCTATAA TGAAGTTAAA CAAGCTGCAA CAGCTAGAAA AGCTCAAAAT	4560
	GCTACAGTTT CAAATGCAAC AAATGAAGAA GTAGCAGAAG CTGATGCAGC AGTAGATGCA	4620
45	GCTCAAAAGC AAGGTTTACA TGACATCCAA GTTGTTAAAT CAAAACAGGA AGTTGCTGAT	4680
	ACAAATCAA AAGTATTAGA TAAATCAAT GCAATTCAAA CACAAGCAAA AGTTAAACCT	4740
	GCAGCTGATA CGGAAGTAGA AAACGCATAT AATACACGTA AACAAGAAAT TCAAAATAGC	4800
50	AATGCTTCAA CTACAGAAGA AAAACAAGCT GCATATACAG AATTAGATAC TAAAAAGCAA	4860
	GAAGCAAGAA CAAATCTTGA TGCTGCAAAT ACAAACAGTG ATGTAACAAC AGCTAAAGAC	4920

55

GCGGAAATCG CTCAAAAAGC AAGTGAACGT AAAACAGCAA TTGAAGCAAT GAATGATTTCG 5040  
 ACTACTGAAG AACACAAGC AGCGAAAGAC AAAGTGGATC AAGCAGTAGT TACTGCAAAC 5100  
 5 GCTGATATAG ATAATGCTGC AGCAAACAAT GATGTGGATA ATGCAAAAAC TACAAATGAA 5160  
 GCTACAATCG CAGCCATTAC ACCTGATGCA AATGTTAAAC CAGCAGCAAA ACAAGCAATT 5220  
 GCAGATAAAG TACAAGCTCA AGAAACAGCA ATTGATGGAA ATAACGGCTC AACAACTGAA 5280  
 10 GAAAAAGCAG CTGCTAAACA ACAAGTTCAA ACTGAAAAAA CAACAGCTGA TGCCGCAATA 5340  
 GATGCAGCAC ATACAAATGC GGAAGTTGAA GCGGCTAAAA AAGCAGCAAT TGCTAAAATT 5400  
 GAAGCGATTG AGCCAGCAAC AACAACTAAA GATAATGCGA AAGAAGCAAT TGCTACGAAA 5460  
 15 GCGAATGAAC GTAAACAGC AATCGCTCAA ACGCAAGACA TTACTGCTGA AGAAATTGCA 5520  
 GCGGCTAATG CGGACGTAGA TAATGCTGTG ACACAAGCAA ATAGCAACAT TGAAGCTGCT 5580  
 AATAGTCAAA ATGATGTAGA CCAAGCGAAA ACGACAGGTG AAAATAGTAT TGATCAAGTA 5640  
 20 ACACCAACAG TTAATAAAAA AGCAACTGCA CGTAATGAAA TCACAGCAAT TTAAATAAC 5700  
 AAATTGCAAG AGATTCAAGC TACGCCAGAT GCAACAGATG AAGAAAAACA AGCAGCTGAT 5760  
 25 GCTGAAGCAA ATACTGAAAA TGGTAAAGCA AATCAAGCCA TTTCAGCAGC AACTACTAAC 5820  
 GCACAAGTTG ATGAAGCTAA AGCAAATGCA GAAGCAGCGA TTAATGCGGT AACACCAAAA 5880  
 GTTGTGAAGA AACAAAGCGC TAAAGATGAA ATTGATCAAT TACAAGCAAC GCAAACAAAT 5940  
 30 GTTATCAATA ATGATCAGAA CGCTACAACA GAAGAAAAAG AAGCAGCTAT TCAACAATTA 6000  
 GCAACAGCAG TTACAGACGC GAAAAATAAT ATTACAGCTG CAACTGATGA TAATGGTGTA 6060  
 GATCAGGCGA AAGACGCTGG AAAGAATTCA ATTCAAAGCA CGCAACCAGC AACAGCGGTT 6120  
 35 AAATCAAATG CTAAAAATGA TGTGATCAA GCTGTGACAA CTCAAAATCA AGCAATTGAT 6180  
 AATAEAACTG GTGCTACAAC TGAAGAGAAA AATGCAGCAA AAGATTTAGT TTAAAGCT 6240  
 AAAGAAAAAG CGTATCAAGA TATCTTAAAT GCACAAACAA CTAATGATGT TACGCAAATT 6300  
 40 AAAGATCAAG CAGTTGCTGA TATTCAAGGT ATTACTGCAG ATACAACAAT TAAAGATGTT 6360  
 GCGAAAGATG AATTAGCAAC AAAAGCAAAC GAACAAAAAG CGTTATTGC ACAAACTGCA 6420  
 GATGCGACTA CTGAAGAAAA AGAACAAGCA AATCAACAAG TAGACGCACA ATTAACACAA 6480  
 45 GGTAAATCAA ATATTGAAAA TGCACAGTCA ATCGATGATG TAAACACTGC AAAAGATAAT 6540  
 GCAATTCAAG CAATTGACCC AATTCAAGCA TCAACAGATG TTAACCGAA TGCAAGAGCG 6600  
 50 GAATTGCTAA CTGAAATGCA AAATAAAATA ACTGAAATAC TTAATAATAA TGAGACTACT 6660  
 AATGAAGAAA AAGGTAACGA TATTGGACCA GTTAGAGCAG CATATGAAGA AGGTTTAAAT 6720

55

AAAGTTCAAC AACTTCATGC AAATCCTGTT AAGAAACCAG CAGGTAAAAA AGAATTAGAT 6840  
 CAAGCTGCAG CTGATAAGAA AACACAAATA GAACAAACAC CAAATGCATC ACAACAAGAA 6900  
 5 ATTAATGATG CAAAACAAGA AGTTGATACT GAATTAAATC AAGCGAAAAAC AAATGTCGAT 6960  
 CAATCATCAA CAAATGAATA TGTGATAAT GCAGTTAAAG AAGGAAAAGC TAAAATTAAT 7020  
 GCAGTTAAAA CATTTAGTGA GTACAAAAAA GATGCTTTAG CTAAAATTGA AGATGCATAT 7080  
 10 AATGCTAAAG TAAACGAAGC GGATAACTCT AACGCATCGA CTTCAAGTGA AATTGCTGAA 7140  
 GCGAAACAAA AACTTGCTGA ATTAAAACAA ACTGCGGATC AAAATGTAA TCAAGCTACT 7200  
 TCTAAAGATG ACATTGAAGT TCAAATTCAT AATGACTTAG ATAATATTAA CGATTACACA 7260  
 15 ATTCCAACAG GTAAAAAGA ATCAGCTACA ACAGATTTAT ATGCTTATGC AGATCAGAAG 7320  
 AAAAATAATA TTTCAGCTGA CACTAATGCA ACACAAGATG AAAAGCAACA AGCAATTAAG 7380  
 CAAGTTGACC AAAATGTTCA AACTGCATTA GAAAGCATTA ATAATGGTGT GGATAATGGT 7440  
 20 GACGTTGATG ATGCATTAAC ACAAGGTAAG GCAGCAATTG ATGCTATTCA AGTAGATGCT 7500  
 ACTGTTAAAC CTAAAGCGAA CCAAGCTATT GAAGTTAAAG CAGAAGATAC GAAAGAATCT 7560  
 25 ATTGATCAAA GTGACCAGTT AACTGCTGAA GAAAAAAGT AAGCATTAGC AATGATTAAA 7620  
 CAAATTACAG ATCAAGCTAA ACAAGGTATT ACTGATGCAA CAACAACTGC TGAAGTTGAA 7680  
 AAAGCGAAAG cTcAAGGACT TGAAGCATTT GATAACATTC AAATCGACTC AACAGAAAAA 7740  
 30 CAAAAAGCTA TCGAAGAATT AGAAACTGCA CTAGACCAGA TTGAAGCAGG TGTAATGTC 7800  
 AACGCTGATG CTACAACTGA AGAAAAAGAA GCGTTTACGA ATGCTTTAGA AGACATTTTA 7860  
 TCAAAAGCAA CTGaAGATAT TTCTGATCAA ACTACAAATG CAGAAATCGC TACTGTCAA 7920  
 35 AATAGTGC GC TTGAACAACT TAAAGCACAA CGTATTAATC CTGAAGTTAA GAAAAATGCT 7980  
 TTGGAAGCAA TCAGAGAAGT GGTTAACAAG CAAATAGGAA cAATTAAAAA TGCAGATGCA 8040  
 GATGCATCGG CGGAAAGAA TTGCACGTAC GGGATTTAGG TAGATATTTT GGACCGATTT 8100  
 40 GCTGGATAAA TTAGGGTAA AACCCCAACC AATGCCGAAG TTGCCTGAAT TACCA 8155

(2) INFORMATION FOR SEQ ID NO: 64:

45 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1630 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

55

CTGTTTTATT TGCAGCACCC ATACTGGAAA TCACTTTAAT CCCTCGGTCA AGACACTCTT 120  
 TCATTAAGTG TACTTTGTAC ATTATTGTAT CACTTGCATC TACAAAATAA TCTATATCGT 180  
 5 AGTTATCGAA AATTCTCTCA TATGTCTCTT CTGTATAAAA CATATGTAAG GGCGTGACTT 240  
 TACAATCTGG ATTAATTAAT TTAATACGTT CTTCATCAA AGAAACTTTA CTTTGTCTTA 300  
 CCGTTGTAGT TAAAGCGTGT AATTGTCTGT TTACATTTGT AATATCAACA TCATCTTTAT 360  
 10 CTATTAATAT AATATGACCA ATATTGCTTC TTGCTAATGC TTCAGCAGCA AATGAACCAA 420  
 CACCTCCAAC GCCAAGTATG ACAACAGTTT GTTGCTTCAA TAAATCTAAA CCTGTGTGTC 480  
 CAATCGCTAG TTCATTTCTT GAAAATTGAT GTTTCATTAT TTTACCTCTT TCACTGATTT 540  
 15 ATACATAAGT ACATAGTAAC TTAAATTTT ATATTTAGCA TTATCACTTT GATTATTTTC 600  
 CCAAATTC ACGAGGAAAC ATTTATTTAA CGCTATAAAA CCCAACTAAT TCTTTATTAA 660  
 AAACCTTAAAG AAACGCATAA AAATACGCAA GACAAAGTCT TGCGTATCGA TAGAGTCCGT 720  
 20 ATTGCCGTAG TTATAATAGC TTGATCATTC GGCCTGTTAT ATACAGGTGG GTGCCCTGTT 780  
 TCTTGTTTTG TACGTCCTTC ATATAAGGCG TGTACGCTGC AAGAAAACCC ATTGGGCTCC 840  
 25 CTTGATCAAA GAGTGTAGG CCCAAATTAA AAAGCAAAC TACGAACAAC TCAGATGACT 900  
 ATCTTATGAT GTTATATTAC CACATAATTA AAATTAATGA AATTATAACA AACCAGGTT 960  
 TATTGATTTT TTAAATTTA GTGACGAATT CGCAAAGAAA GTTCTTCTAA TTGTTTATCA 1020  
 30 GAAACTTCAC TAGGCGCATT CGTTAATAAA CATGTAGCAG ATGCTGTTTT AGGGAATGCG 1080  
 ATTGTATCTC TCAAGTTTGT TCTATTAGTC AATAACATGA CTAATCGGTC TAATCCTAAT 1140  
 GCAATACCGC CATGTGGTGG TGCACCATAT TTAAATGCAT CTAGTaAGAA GCCGAAGTGT 1200  
 35 TCCTgTGCTT GTTCTTTAGT AAATCCAAGA ACTTCGAACA TTTTTTCTTG TAACTCACCA 1260  
 TCATGAATTC TGATTGAACC GCCACCTAAT TCATAACCAT TTAATACTAT GTCATAAGCA 1320  
 TTTGCCTCAG CTTCTTCTGG CGCAGTGCCA AGCTTAGCAA TATCAGCTTC TTTTGGAGAT 1380  
 40 GTAAATGGAT GATGTGCTGC AACGTAACGT TTCGCATCTT CATCATATTC TAATAATGGC 1440  
 CAATCTGTCA CCCATAAGAA GTTTAATTTT GTTTCATCGA TTAAACCTAA TTCTTTAGCT 1500  
 AATTTGACAC GTAATGCACC TAACTTTGT GCAACGACAT TTGGTttGTC TGCAACAAAC 1560  
 45 ATTACTAAGT CACCAGCTTC AGCACCAGTT AATGTAAGTA ATGTTTCAAC ATTTTCTGTT 1620  
 cAAAGAAACG 1630

50 (2) INFORMATION FOR SEQ ID NO: 65:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 732 base pairs

55

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

	CAATTGGACA TCTTGATGA AAAGGACAAC CTTGCGGCGG ATTACTTGGC GAAGGTAATT	60
10	CTCCTTTTAA TATAATTCTA TTGTTATTAT GTTTATCAAT TTGTGGTATT GATGAAATCA	120
	ACGCTTTTGT ATATGGATGT TTGGGATTTT CATAAATTTC TTTATCAGAT GCGATTTCAA	180
	CTATATGACC TAAATACATA ACTCCAATGA CATCACTTAT ATGTTTTACT ACACTTAAAT	240
15	CATGTGCGAT AAATAAATAG CTTAAGTTAA ATTGTTCTTG TAAATCTTTT AATAAATTCA	300
	GTA CTTGAGA TTGAACAGAT ACATCTAATG CACTTACAGG CTCATCAGCA ACAATTAAAC	360
	TGGGACGCAA AGCCAATGCT CTTGCAATTC CCACTCTTTG TCTCTGTCCA CCTGAAAATT	420
20	CATGTGCATA TTtATAATAT GCATCTTCAC TTAGGCCAAC ACATTTTAAT AAATATAGTA	480
	CTTCTTTTTT TATTTCTTCT TTTGGCAATT TTTTATAATT TAAAATAGGT TCTGAAATGA	540
	TATCTCCAAC CATTTGCATC GGATTCAATG ATGCATACGG ATCTTGAAAT ATCATCTGAT	600
25	ATTGTTGTCG TGATTTTCTG AGTTTTTTTAC CTTGTAATCT TGTtATATCT TCACCATTAA	660
	CAATTATTGA GCCTGAAGTT GCATCTTCAA GCCTGATAAT CACTTTACCT AACGTTGACT	720
30	TACCACAACC CG	732

(2) INFORMATION FOR SEQ ID NO: 66:

(i) SEQUENCE CHARACTERISTICS:

35	(A) LENGTH: 5838 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

	AATATATTCA TATGTTTCAT CAACAATATT AGCTGCTTTT TGAATTAAAG CAATTTGTC	60
	AGCATCTTTG ACGTCTCTAA TTTTATCTAC AGTATTAGAA ATGCTTATTA ATGATATACG	120
45	GCTTTTATTT AATTCAAGGT ATGTATCATA ACTTACATGA TGCCCTCAA AACCTACATT	180
	TTCAAAATTT TCTTGGTGTA GCAATTCTTT AATCTCACCA ATAATAGTAG ATTTACGATT	240
50	AATAATTTCA TAATTTGGCG CCTGCTTAGT TGCTTGATCA ATATATCTAA AGTCTGTTAT	300
	CAAATATTGT TTATCTTTAG ATATGATAAG TGCTCCACTG GTACCAGTAA AACCTGATAA	360
	ATATCTTCTA TTGTAATCCG AAAGAATGaT AATCGCATCT AAATGTTTTT GTTCTAAAAAT	420

55

	CAACTTTATA CATTAAAATA ATATCATAAT AAGGATAAAA AATAATAGAT ATTGATTTTA	540
	GGGAGATAGT AATGAAAAAA TTGGTTTCAA TTGTTGGCGC AACATTATTG TTAGCTGGAT	600
5	GTGGATCACA AAATTTAGCA CCATTAGAAG AnAAAACAAC AGATTTAAGA GAAGATAATC	660
	ATCAACTCAA ACTAGATATT CAAGAACTTA ATCAACAAAT TAGTGATTCT AAATCTAAAA	720
	TTAAAGGGCT TGAAAAGGAT AAAGAAAACA GTAAAAAAC TGCATCTAAT AATACGAAAA	780
10	TTAAATTGAT GAATGTTACA TCAACATACT ACGACAAAGT TGCTAAAGCT TTGAAATCCT	840
	ATAACGATAT TGAGAAAGAT GTAAGTAAAA ACAAAGGCGA TAAGAATGTT CAATCGAAAT	900
15	TAAATCAAAT TTCTAATGAT ATTCAAAGTG CTCACACTTC ATACAAAGAT GCTATCGATG	960
	GTTTATCACT TAGTGATGAT GATAAAAAAA CGTCTAAAAA TATCGATAAA TTAAACTCTG	1020
	ATTTGAATCA TGCATTTGAT GATATTAAAA ATGGCTATCA AAATAAAGAT AAAAAACAAC	1080
20	TTACAAAAGG ACAACAAGCG TTGTCAAAAT TAAACTTAAA TGCAAAATCA TGATAGGAGT	1140
	CTTTTAATGC GTAATATAAT ATTTTATCTT GTACTTATTA TTGCTGCGAT TGGATTAGTA	1200
	ATGAATCTAG ATGCCTTTAT TTTTCAATC GTCAGAATGT TAATCAGCTT TGcgTAaTAG	1260
25	CTGGTATTAT TTATCTGATT TATTATTTCT TCATCTTAAC TGAAGACCAA CGCAAATATC	1320
	GCAAAGCAAT GCgTrAaGTA TAAAGAAAT CAAAGAAGAA AATAGATAAA AAAACGGAAG	1380
	CACTTGTAGG TAAAATAGTC TACGTGCTTC CATTTTTTAT TCTAAAAACT ACTTTCTAAA	1440
30	CATCCATTCA TCTGAACGAT ATTTTTCAGT TAATTCTTCC ACTTCTGCCA ATTGAGCTTC	1500
	TGcTAATTCA AGTGGCTTTA ATTCTATATT TAAACCTTTC TTA AACCTT TCTCGAAAGC	1560
	TTCTTCCATT TGAATAATAG TAATGTGTTC ATCTGAAATA TCATTGATGG CAACTGCTTT	1620
35	TTCAACGAAT GCCTCTTTCA TTTTAAATTT TAATCTTTCA TTTTATAAA TraACATATC	1680
	AAACAGTTCA TCAATATCAA TATCTTGTA AATCGAACCG TGTGGAGGA TTACGCCCTT	1740
40	TTGTCTCGTT TGAGCACTCC CAGCAATCTT ACGGCCTTCA ACAACTAGCT CATAACCACT	1800
	TGGTGCATCA AAACACACTG AACTTCGAGG TTGTTTAAAT TTTTGACGCT CTTCAGGCGT	1860
	TTIAGGTACC GCAAAATAAG TATCAAATCC TAAGTTTTTA AATCCTTCTA ATAATCCTTG	1920
45	TGAAATCACT CTGTACGCTT CTGTAAGTGT AGAAGGCATA TTCGGATGCG ATTCAGGCAC	1980
	AATCACACTG TAAGTTAACT CTTTATCATG TAGCACCCCA CGGCCACCAG TTTGACGCCT	2040
	TACGAGACCA AAACCTTTCT CTTTAACTT ATCAATATCA ATTTCTTTTT GTAGCCTTTG	2100
50	GAAATACCCT ATTGATAATG TTGCAGGATT CCATGTGTAA AAACGTATAA CTGGATCAAT	2160
	TTACCTCTA GAGACAAAAT TTAATAACGC TTCATCCATT GCCATATTAT AATATGGGTC	2220
55		

	AAATGTATAA TATTTGATTC GCTAATTAAT CAATTTAACT AAATGAATAA TAATTGCAAT	2340
	TCTTTAGTGA AATATTTTGA TAATTTGACC TAACAGTCTT ATAATTATAT TATCGTTTAA	2400
5	TTAGGGAGGA TGCAAGATGA GTGCTAGTTT GTACATCGCA ATAATTTTAG TTATAGCAAT	2460
	TATTGCTTAT ATGATTGTTC AACAAATTCT TAACAAGCGA GCTGTTAAAG AATTAGATCA	2520
	AAATGAATTC CATAATGGGA TTAGAAAAGC TCAAGTCATC SATGTTAGAG AGAAAGTTGA	2580
10	CTATGACTAC GGTACATTA ATGGGTCTCG CAATATTCCT ATGACAATGT TCAGGCAACG	2640
	ATTCCAAGGA TTAAGAAAAG ATCAACCGGT ATACTTATGT GATGCCAATG GGATTGCTAG	2700
15	CTATAGAGCC GCTCGTATTT TGAAAAAGAA TGGATATACA GATATCTATA TGTTAAAAGG	2760
	CGGCTATAAA AAATGGACTG GAAAAATAAA GTCTAAAAAA TAGTTTTTGT AAATTTAATA	2820
	TACGATTTAA TAAATCTGA GTGTAAATTG ATCATCAATA ACAATACTCA GATTTTAATT	2880
20	TTTTAACAAA GTCTGTTACT ATATTTCTCT AGCTTCACTG ATCATTAAAC TTAGTTTCAG	2940
	CATAATAAAG AAAGTTCAGC TCATTTTCAA TACGATTCAA TTACCGCAAT CTAAAAAATG	3000
	AAAAGACAAT TTCTATGAAA GAATAATACC AAACCCTAAG AGTTATTACT TCGGTTTAGT	3060
25	TTTCTTGTTT AAATAGAAAT TGTCTTTTTC AATTGATTTT GAAACCATTA TCCTTAAATC	3120
	TTCATACAAA GTTAGAATAA TAATTCTCGG AATATGTGTT TAATACTTTA TTTTTCCTGT	3180
	TTAAGATTTT CAAACTTTAA TATTGGTTTA CGAGCAGCTG TAGCTTCGTC TAATCGATCA	3240
30	ATCACAGTTG TATGTGGTGC TTCTAGCact TTATCAGGAT CATTTTTAGC TTCTTCAGCA	3300
	ATACTAATTA ATGTATCGAT AAAATAATCA AGTGTCTTCT TAGACTCTGT CTCAGTCGGT	3360
35	TCAATCATCA TACCTTCTTC AACATTTAAT GGGAAGTATA TTGTTGGTGG ATGTACACCG	3420
	AAATCTAATA ATCGCTTAGC CATGTCTAAA GTACGTACAC CAAATTCTTT TTGACGCACA	3480
	CCACTTAACA CAAACTCGTG TTTACAATAT TGTTTATAAG GTATTTCAA GTGTTTAGAT	3540
40	AAACGTGCTT TAATATAATT CGCATTAAGA ACCGCTGCTT CAGAAACCTC TTAAAGTCCA	3600
	GTTGCTCCCA TAGTTCGAAT ATACGTATAA GCTCTTAAGT AAATACCAA GTTACCATAA	3660
	AATGGTTTTA CACGTCCGAT AGAATTTTTA ATGTCATTAT CATATTTAAA TTTGTCGCCA	3720
45	TCTTTAATAA CCATTGGCTT TGGTAAGTAA CTTGCTAGTT CTTTACTAC ACCGACTGGA	3780
	CCTGAACCAG GACCGCCACC ACCATGTGGA CCAGTAAATG TTTTATGCAA GTTTAAATGA	3840
	ACAGCATCAA ATCCCATATC TCCTGGGCGA ACTTTGTCCA TAATAGCGTT TAAATTCGCA	3900
50	CCATCATAAT ATAATAGACC ACCAGCATTA TGGACGATTT CACGGATTTC CATAATATTT	3960
	TTTTCGAAAA TACCTAAAGT GTTTGGATTA GTTAACATAA TAGCTGCTGT ATTTTCATTT	4020
55		



	GATTTAAATC CTGCAAATGa AGCTGAGGCT GGaTTCGTAC CATGCGCAGA ATCTGGcACA	4140
	ATGACTTCAT CACGATGACC TTCACCATTA TTCTCATGGT AAGCTTTAAA TATCATCAAT	4200
5	GCAGTCCAAT CACCATGTGC GCCAGCAGCT GGTGTGAATG TCACCTCATC CATAACAGTA	4260
	ATTTCTTTTA ATTCTTCTTG CAAACTATAA ATAATTCTA ATGAACCTTG AACTTGATCT	4320
	TCATCTTGTA ATGGATGTGA TTCACTAAAT CCTGGTATTC TAGCAACCTT TTCATTAATT	4380
10	TTAGGGTTAT ACTTCATCGT ACATGAACCC AATGGATAAA ATCCGTTGTC TACACCGAAA	4440
	TTTTTATTTG AAAGTTCAGT ATAATGACGT ACTAAGTCTA GTTCAGCAAC TTCAGGAAAC	4500
15	TCCGCTTTGT TTTTACGAAT AAATTTATCA TCTAACAATG ACTCAACAGA ATTTGTTTTA	4560
	ATATCACTTT TTGGTAATGA ATATGCATAT CTGCCTTCAC GAGATCTTTC AAAAATTAAT	4620
	GGACTTGATT TACTAGTCAT TTAACCTACC AGCCTTTTCT ACAAATGTAT CGATTTTCATC	4680
20	TTTTGTTCTT AATTCAGTTA CAGCTATTAA CATGTGATTT TTAAAGTCGT CTGAAACAAC	4740
	ACCTAAATCA AAACCACCGA TAATATTGTA CTTCACTAAT TCCTCGTTAA CTTGTTGAAT	4800
	TGGTTTGTC AATTTGACTA CAAACTCATT GmnAAGnTGT ACCATCTAAT ACTTCAAAAC	4860
25	CTTTTTTAAT AAATTGTTGT TTAGCATAGT TAGCATGTTT TATATTTTGA ACTGCAATAT	4920
	CATAGATACC TTGTTTACCA AGTGCTGACA TTGCAATTGA TGaCGcTAAA GCATTTAATG	4980
	CTTGTTTGA ACAAATATTA GATGTCGCTT TATCGCGTCG AATATGTTGT TCACGTGCTT	5040
30	GTAATGTAA TACAAAGCCA CGATTACCTT CATCATCTTG TGTTTGACCG ACTAATCTAC	5100
	CTGGCACTTT ACGCATTAAAC TTTTTCGTCG TTGCAAAATA TCCACAATGT GGCCACCGA	5160
35	ATTGAGCAGG AATTCCGAAT GGCTGAGTAT CACCTACAAC AATATCTGCA CCAAATGAAC	5220
	CTGGAGGTGT AAGTAATCCC AATGCTAATG GATTTCATA TACGATAAAT AATGCTTTTT	5280
	TATCTTCAAT AAAGCTATGA ATCTTTTCAA GATCTTCAAT TGAACCGTAA AAGTTTGAT	5340
40	ATTGTACTGC AACAGCTGCT GTTTCATCAT CCACTGCTGC TTCTAATTTT TTCAAATCTG	5400
	TAACAGTGCC ATCTAAATCG ATTTCCACTA CTTTGAATTC CTTACGCGTC TTAGCATAAG	5460
	TATGAAGTAC TTGTAATGCT TGATAATGTA AACCTTTTGA GACTACAATT TTATTTTTCT	5520
45	TTGTTTGACT AAATGCTAAG ATACATGCTT CAGCAAAGCT AGTCATCCCA TCATACATAG	5580
	AAGAATTTGC TACATCCATA TCTGTAAATT CACAAATTAA AGTTTGGAAC TCAAAAATGG	5640
	CTTGTAATTC ACCTTGAGAA ATTTCCGGTT GATATGGCGT ATATGCTGTG TAAAATTCTG	5700
50	ATCTTGAAAT CATAGCATCC ACAACTGATG GCGCGTAATG ATCATAAACA CCAGCACCCA	5760
	rAAATGATGT ATGCGTTTCT TTAGTGATAT tCTTGCTkGC AATGGGGATT TAAACnTCTA	5820
55		

## (2) INFORMATION FOR SEQ ID NO: 67:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18355 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

5	AThATAATTG GCTTTGCTAA TAATTACTTC CCTGAATTAC aAGTATTAGC AAACGAAATA	60
10	AAATCTGATA TGGCTAGTTC ATTAaAACAA TGATATTTT ATTTAAATTT TTaAAGCTTT	120
15	GTACGAAATT GTACAAAGCT TTTTGGTGC GTATTGTATG GGCAACAAC TGACGATGAA	180
20	AATCCGTTAC AGGATTGGTA ATAGGAAATG TTAGCGAAAG ACAAGGGTAT CCATTGTAGA	240
25	TTAACAAAG GACGTTTCCA CAAGTGTGGG TTATTCTCAC TAAAGCAATA CGCAGAGACA	300
30	ACTTACGTAA AATTTTGAAC TGA CTAGAAC GGAAC TCTA CTCAATTATT GATAAAATT	360
35	TTCAAAAGA CTTGAATGTG CTGAGAATAC GAAGTTTATG GAAGGATTAT CAAAATATAA	420
40	ATGTGCATTC ATTTACAACC TTTATTGACA ATGATTCTCA ACTAATATAG TATATAATCA	480
45	AATCGTAATA GTTACGATTT GTTTCTGCA ACTTTTGA AGTTT TAGTT GAGGTGAAAA	540
50	CAATAAAAGC ATCTAAGTGA ATGTAGTTAA CGGACAAC TG CATTGCTTG TAGAGCCACA	600
55	AGAAGCAACT TTAAATAAGG TTTACGGTTG CATTTTGATA CAACAACCGA TTACTAAGTC	660
60	ATGCTTTCCA CTTTGC GGT TAGCATGACT TACCTAATAG ATAGAGCTAT TAGGTT CAGC	720
65	TTCTAAAAA TTACAGTTT AGAGGAATAC AGTTGCTTGc tTCGCAACAA CTGCATAAGA	780
70	GCCATGGTTT TCGCTTTTGC GAATTAGCAT GACTTACCTA CTAGATAGAG CTATTAGGTT	840
75	CATCTTCTAA AAAATTACAG GTTTAGAGGA ATACAGTTGT TTGcTTCGCA ACAACTGCAT	900
80	AAGAGCCTCT AGTAATTAAA ATTACAGAGG CTCTAAAAAT ACATCTAAAG GAGTGTCGTA	960
85	TGAATCGGCA GGTTATAGAA TTTTCTAAGT ATAATCCTTC GGGGAATATG ACGATACTTG	1020
90	TTCATTCAA ACATGATGCT AGTGAATATG CATCTATCGC CAATCAGTTG ATGGCCGCAA	1080
95	CACATGTATG CTGTGAACAG GTAGGCTTTA TAGATCAAC ACAAATGAT GATGGTAATG	1140
100	ATTTTCACTT AGTTATGAGC GGTAATGAAT TTTGCGGTAA TGCGACGATG TCATATATAC	1200
105	ATCATTTGCA GGAAAGTCAT TTGCTTAAAG ACCAACAGTT TAAGGTGAAG GTGTCTGGCT	1260
110	GTTGCGATTT AGTGCAATGC GCAATTCATG ATTGCCAATA CTATGAAGTT CAAATGCCAC	1320
115	AAGCCCATCG TGTTGTGCCA ACAACAATTA ATATGGGTAA TCATTCATGG AAAGCAATAG	1380

	TTCAACATTT	GGTTGAAGCG	TTTGTGCGTG	AgcAACAAATG	GAGTCACAAA	TATAAAACAG	1500
	TAGGTATGAT	GCTTTTTGAT	GAACAACGTC	AATTTTTTACA	GCCATTAATC	TATATACCAG	1560
5	AAATTCAAAG	TTTAATTTGG	GAAAATAGCT	GTGGTTCTGG	TACAgcATCA	ATTGGGGTTT	1620
	TTAATAATTA	TCAACGTAAT	GACGCATGCA	AAGATTTTAC	AGTACATCAG	CCAGGGGGCA	1680
	GTATTTTAGT	GACATCAAAG	CGATGTCATC	AATTGGGATA	TCAAACCTCA	ATTAAAGGAC	1740
10	AGGTTACAAC	TGTAGCTACA	GGaAAAGCAT	ATATAGAATA	AGGAGCCTAC	AATGAATAAC	1800
	TTTAATAATG	AAATCAAATT	GATATTACAA	CAATATTTAG	AAAAGTTTGA	AGCGCATTAC	1860
	GAGCGTGTAT	TACAAGACGA	TCAATATATC	GAAGCATTAG	AAACATTGAT	GGATGACTAT	1920
15	AGTGAATTTA	TTTTAAATCC	TATTTATGAA	CAACAATTTA	ATGCTTGGCG	TGACGTTGAA	1980
	GAAAAAGCAC	AATTaATAAA	ATCACTGCAA	TATATTACAG	CGCAGTGTGT	TAAACAAGTG	2040
20	GAAGTCATTA	GAGCGAGACG	TCTATTAGAC	GGACAGGCGT	CTACCACAGG	TTACTTTGAC	2100
	AATATAGAAC	ATTGTATTGA	TGAAGAGTTT	GGACAATGTA	GTATAGCTAG	CAATGACAAA	2160
	TTATTGTTAG	TTGGTTCAGG	TGCATATCCA	ATGACGTTAA	TTCAAGTAGC	AAAAGAAACA	2220
25	GGTGCTTCAG	TTATCGGTAT	TGATATTGAT	CCACAAGCCG	TTGACCTAGG	GCGCAGAATC	2280
	GTTAACGTCT	TAGCACCAAA	TGAAGATATA	ACAATTACGG	ATCAAAAGGT	ATCTGAACTT	2340
	AAAGATATCA	AAGATGTGAC	GCATATCATA	TTCAGCTCGA	CAATTCCTTT	AAAGTACAGC	2400
30	ATTTTAGAAG	AATTATATGA	TTTAACAAAT	GAAAATGTCG	TAGTTGCAAT	GCGCTTTGGT	2460
	GATGGCATCA	AAGCAATATT	TAATTATCCG	TCACAAGAAA	CAGCGGAAGA	TAAGTGCCAA	2520
	TGTGTGAATA	AACATATGAG	ACCACAGCAA	ATTTTTGATA	TAGCACTTTA	TAAAAAAGCA	2580
35	GCTATAAAGG	TAGGTATTAC	GGATGTCTAA	ATTATTAATG	ATAGGCACTG	GTCCgGTCCG	2640
	AATGCAATTA	GCGAATATTT	GCTATTTAAA	ATCAGATTAT	GAGATTGATA	TGGTTGGACG	2700
40	TGCCTCAACA	TCAGAAAAAT	CAAAACGCTT	ATATCAAGCG	TATAAAAAAG	AGAAACAATT	2760
	TGAAGTCAAA	ATACAAAACG	AGGCGCATCA	ACATCTGGAA	GGTAAGTTTG	AAATTAATCG	2820
	TTTGTATAAA	GATGTTAAAA	ACGTTAAGGG	TGAATACGAA	ACGGTTGTCA	TGGCATGCAC	2880
45	AGCAGATGCT	TATTATGACA	CACTACAGCA	ATTGTCGTTA	GAACTTTTGC	AAAGTGTCAA	2940
	ACATGTCATT	TTAATATCAC	CGACATTTGG	TTCGCAAATG	ATTGTCGAAC	AATTTATGTC	3000
	TAAATTTAAT	AAAGATATCG	AAGTGATTTC	ATTCTCAACT	TATCTTGGCG	ATACACGTAT	3060
50	TGTTGATAAA	GAAGCGCCTA	ATCATGTGTT	GACAACAGGT	GTAAAAAAGA	AATTGTACAT	3120
	GGGATCGACA	CATTCAAAC	CAACAATGTG	TCAACGAATC	TCTGCTTTAG	CTGAGCAATT	3180
55							

	TTATGTGCAC CCACCACTAT TTATGAATGA CTTTTCATTG AAAGCCATTT TCGAAGGAAC	3300
	AGATGTACCG GTTTATGTGT ATAAGTTATT TCCTGAAGGA CCGATAACGA TGACACTAAT	3360
5	CCGTGAAATG CGTTTAATGT GGAAGGAAAT GATGGTTATT TTACAAGCAT TTAGAGTGCC	3420
	GTCAGTCAAC CTGCTTCAAT TTATGGTGAA GGAAAATTAT CCAGTACGTC CTGAAACTTT	3480
	GGATGAAGGT GATATTGAGC ATTTGCGAAAT CTTGCCAGAT ATCTTACAAG AATATCTGCT	3540
10	TTATGTAAGA TATACCGCAA TCCTCATTGA TCCATTTTCA CAGCCAGACG AAAACGGACA	3600
	TTACTTTGAT TTTTCAGCTG TACCATTAA GCAAGTCTAT AAAAATGAAC AGGATGTTGT	3660
15	TCAAATTCCA AGAATGCCAA GTGAAGATTA TTACAGAACG GCGATGATTC AGCATATTGG	3720
	GAAAATGCTA GGTATCAAAA CGCCAATGAT TGATCAGTTC CTAACCTCGCT ATGAAGCAAG	3780
	TTGCCAGGCG TACAAGGATA TGCATCAAGA TCAACACTTA TCTTCTCAAT TTAATACAAA	3840
20	TCTATTTGAA GGAGATAAAG CACTCGTCAC AAAATTTTGG GAAATCAATA GAACGCTTTC	3900
	ATAATAAGGG TTTGAAGTTT TATAATAGAA AAAAATTATT GAATTATGTT TGACATTTAC	3960
	ATAAAAATAA GCAAATAATT GAGAAAAATA ATCATTACGA TTTGATTAAG TAATGCAACT	4020
25	TATCAATTTA GAAAGAGGAA AAGCAAATGA GAAAATAAC TAAAATGAGT GCAATGTTAC	4080
	TTGCATCAGG GCTAATTTTA ACTGGTTGTG GCGGTAATAA AGGTTTAGAG GAGAAAAAAG	4140
	AAAACAAGCA ATTAACGTAT ACGACGGTTA AAGATATCGG TGATATGAAT CCGCATGTTT	4200
30	ACGGTGGATC AATGTCTGCT GAAAGTATGA TATACGAGCC GCTTGACGT AACACGAAAG	4260
	ATGGTATTAA GCCTTTACTA GCTAAAAAGT GGGATGTGTC TGAAGATGGG AAGACATACA	4320
35	CGTTCCATTT GAGAGATGAC GTTAAATTCC ATGATGGTAC GCCATTTGca TGctGACGCA	4380
	GTTAAGAAAA ATATTGACGC AgTTCAAGAA AACAAAAAAT TGCATTCTTG GTTAAAGATT	4440
	TCGACATTAA TTGACAATGT TAAAGTTAAA GATAAGTACA CGGTTGAATT GAATTTGAAA	4500
40	GAAGCATATC AACCTGCATT GGCTGAATTA GCGATGCCTC GTCCATATGT ATTTGTGTCT	4560
	CCAAAAGACT TTaAAAACGG TACAACAAAA GATGGCGTTA AAAAGTTCGA TGGTACTGGT	4620
	CCATTTAAAT TAGGTGAACA CAAAAAGAT GAGTCTGCAG ACTTTAACA AAATGATCAA	4680
45	TACTGGGGCG AAAAGCTAA ACTTAACAAA GTACAAGCAA AAGTAATGCC TGCTGGTGAA	4740
	ACAGCATTCC TATCAATGAA AAAAGGTGAA ACGAACTTTG CCTTCACAGA TGATAGAGGT	4800
	ACAGATAGCT TAGACAAAGA CTCTTTAAAA CAATTGAAAG ATACAGGTGA CTATCAAGTT	4860
50	AAGCGTAGTC AACCTATGAA TACGAAAATG TTAGTTGTCA ATTCTGGTAA AAAAGATAAC	4920
	GCTGTGAGTG ACAAACAGT CAGACAAGCG ATTGGTCATA TGGTAAACAG AGATAAAATT	4980
55		

	ACAGACATTA ATTCGATAT GCCAACACGT AAGTATGACC TTAAAAAGC AGAATCATTA	5100
	TTAGATGAAG CTGGTTGGAA GAAAGGTAAA GACAGCGATG TTCGTCAAAA AGATGGTAAA	5160
5	AACCTTGAAA TGGCAATGTA CTATGACAAA GGTTCTTCAA GTCAAAAAGA ACAAGCAGAA	5220
	TACTTACAAG CAGAAATTAA GAAAATGGGT ATTAAGTTAA ACATCAATGG CGAAACATCA	5280
	GATAAAATTG CTGAACGTCG TACTTCTGGT GATTATGACT TAATGTTCAA CCAAACTTGG	5340
10	GGATTATTGT ACGATCCACA AAGTACTATT GCAGCATTTA AAGAGAAAAA TGGTTATGAA	5400
	AGTGCAACAT CAGGCATTGA GAACAAAGAT AAAATATACA ACAGCATTGA TGACGCATTT	5460
	AAAATCCAAA ACGGTAAAGA GCGTTCAGAC GCTTATAAAA ACATTTTGAA ACAAATTGAT	5520
15	GATGAAGGTA TCTTTATCCC TATTTACAC GGTAGTATGA CAGTTGTTGC ACCaAAAGAT	5580
	TTAGAAAAAG TATCATTCAC ACAATCACAG TATGAATTAC CATTCAATGA AATGCAGTAT	5640
	AAATAAAGGA GCAATTAGAT GTTCAAATTT ATCTTAAAC GTATTGCGCT CATGTTTCCA	5700
20	TTGATGATTG TAGTAAGTTT TATGACATTT CTATTGACGT ATATTACAAA TGAAAATCCA	5760
	GCTGTGACAA TTTTACATGC ACAAGGGACG CCAAAATGTAA CACCAGAGTT GATTGCAGAA	5820
25	ACGAATGAGA AGTACGGTTT CAATGATCCA TTATTAATTC AATATAAAAA TTGGTTACTT	5880
	GAAGCGATGC AATTTAATTT TGGTACAAGC TACATTACAG GTGACCCAGT TGCTGAACGT	5940
	ATTGGTCCAG CATTTATGAA TACATTGAAA TTAACAATAA TTTCAAGTGT TATGGTGATG	6000
30	ATTACATCAA TTATTTTAGG TGTAGTTAGT GCATTAAAAA GAGGAAAGTT CACTGATCGT	6060
	GCGATACGTT CAGTGGCTTT CTTTCTAACT GCATTACCAT CATATTGGAT AGCTTCAATA	6120
	CTTATTATTT ACGTTTCAGT GAAGTTAAAC ATATTGCCGA CTTCTGGATT AACAGGTCCA	6180
35	GAAAGTTACA TATTGCCAGT GATCGTTATT ACGATTGCCT ATGCTGGTAT TTACTTTAGA	6240
	AATGTTAGAC GCTCGATGGT GGAACAATTA AATGAAGATT ATGTACTTTA TTTAAGAGCA	6300
	AGCGGTGTGA AATCTATCAC ATTAATGTTG CATGTGTTGC GTAATGCTTT ACAAGTTGCG	6360
40	GTATCAATCT TTTGTATGTC TATACCAATG ATAATGGGTG GACTAGTTGT TATCGAGTAT	6420
	ATCTTTGCAT GGCCTGGACT AGGTCAATTA AGTTTAAAAG CAATACTTGA ACACGATTTT	6480
45	CCAGTCATTC AAGCATATGT ATTAATTGTA GCGGTATTAT TTATTGTATT TAATACATTA	6540
	GCAGATATCA TTAATGCGCT ATTAAATCCA AGATTAAGGG aGGGCGCACG ATGATAATTT	6600
	TAAAmCGATT ATTMCArGWT AAAGGTGCAG TAATTGCTTT AGGCATTATT GTATTATATG	6660
50	TCTTTTTAGG ATTAGCAGCA CCACTTGTGA CATTTTATGA TCCTAACCAT ATCGATACAG	6720
	CAAACAAATT TGCTGGCATG AGTTTTCAAC ATCTACTAGG TACTGACCAT TTAGGTAGAG	6780
55		

	TATTTGTTTC TGTACTTATT GGATCTATTT TAGGATTCTT ATCAGGATAT TTCCAAGGGT	6900
	TTGTTGACGC CTTAATCATG CGTGCGTGTG ATGTTATGTT GGCATTCCCA AGTTATGTTG	6960
5	TAACGTTAGC ATTAATTGCA TTGTTTGGAA TGGGTGCCGA AAATATTATC ATGGCAITTA	7020
	TTTTGACGCG TTGGGCATGG TTCTGTCGTG TTATACGTAC AAGTGTATG CAGTACACTG	7080
	CTTCTGACCA TGTAAGATTT GCTAAAACAA TCGGTATGAA TGATATGAAA ATTATTACCA	7140
10	AACATATTAT GCCATTAACA TTAGCAGATA TTGCTATCAT CTCTAGTAGC TCGATGTGTT	7200
	CAATGATCTT GCAAATATCT GGCTTTTCAT TTTTAGGATT AGGTGTCAAA GCGCCTACTG	7260
	CAGAGTGGGG CATGATGCTT AACGAaGCTA GAAAAGTGAT GTTTACACAT CCTGAAATGA	7320
15	TGTTTGCGCC AGGTATTGCC ATAGTGATTA TAGTGATGGC ATTTAACTTC TTATCCGATG	7380
	CTTTACAAAT TGCTATTGAT CCCC GCATCT CTTCTAAAGA TAACTTCGT TCTGTGAAAA	7440
20	AAGGAGTGGT GCAATCATGA CATTGTTAAC AGTTAAACAT TTGACGATTA CAGATACCTG	7500
	GACAGATCAA CCACTCGTGA GTGATGTGAA TTTTACATTA ACTAAGGGTG AAaCTTTAGG	7560
	CGTTATTGGA GAAAGTGGTA GTGGTAAATC AATCACTTGT AAATCGATTA TTGGTTTGAA	7620
25	TCCCCGAACGA CTCGGGGTGA CAGGTGAAAT TATCTTTGAT GGTACaCAA TGTTGTCATT	7680
	ATCTGAATCG CAATTGAAAA AGTACCGTGG TAAAGACATT GCGATGGTCA TGCAACAAGG	7740
	TAGTCGTGCC TTTGACCCAT CAACTACTGT CGGTAAACAA ATGTTTGAGA CTATGAAAGT	7800
30	ACATACGTCA ATGTCTACAC AAGAAATTGA AAAGACATTG ATTGAATATA TGGATTATTT	7860
	AAGTTTGAAA GATCCTAAAC GTATATTAAA ATCATACCCT TACATGTTAT CAGGAGGAAT	7920
	GTTACAGCGA TTGATGATTG CTTTAGCGTT AgcTTTgAAA CCAAAGTTAA TCATTGCTGA	7980
35	TGAGCCGACA ACGGCTTTAG ATACAATTAC ACAATATGAT GTACTGGAAG CATTTATAGA	8040
	TATTAAAAAA CACTTTGACT GTGCGATGAT TTTCAATTCA CATGATTTAA CGGTTATTAA	8100
	CAAGATTGCA GACCGTGTTG TTGTGATGAA AAATGGTCAG CTTATTGAAC AAGGGACACG	8160
40	TGAATCAGTC TTGCATCATC CAGAACATGT TTATACGArt ATTKTATTAT CAACGAAGAA	8220
	GAAGATTAAT GATCATTTTA AACATGTGAT GAGGGGTGAT GTACATGATT AAAATTAAAG	8280
45	ATGTTGAAAA GTCATATCAA AGCGCACATG TTTTAAAGCG TCGTCGAACA CCTATCGTGA	8340
	AAGGTGTGTC ATTTGAGTGT CCAATCGGTG CGACGATTGC GATTATCGGA GAAAGTGGTA	8400
	GCGGTAAATC GACGTTGAGT CktATGATAT TAGGTATTGA GAAACCGGAT AAAGGTTGTG	8460
50	TAACCTTAAA TGATCAACCG ATGCATAAGA AGAAAGTGAG ACGTCATCAA ATTGGTGCTG	8520
	TATTTCAAGA TTATACGTCA TCATTACATC CATTTAGAGC TGTTAGAGAA ATCTTATTTG	8580
55		

	TGTTGGAAGA AGTCGGTCTA TCTAAGGCAT ACATGGATAA ATATCCTAAT ATGTTATCAG	8700
	GTGGAGAGGC GCAACGTGTT GCGATTGCGC GTGCAATATG TATTAACCCCT AAATATATTT	8760
5	TGTTTGATGA AGCCATTAGT TCACCTCGACA TGTCAATTCA AACACAAATA TTAGATTTAT	8820
	TGATTCATTT ACGTGAAACG CGTCAGTTGA GTTATATTTT TATCACACAT GATATTCAAG	8880
	CTGCCACGTA TTTATGTGAT CAATTAATTA TTTTAAAAA CGGAAAAATA GAAGAACAAA	8940
10	TTCCGACAAG CGCATTGCAT AAAAGTGACA ATGCTTATAC AAGAGAATTA ATAGAAAAAC	9000
	AACTATCATT CTAAGGAGTG AGATAATGAA AGGTGCAATG GCTTGGCCCT TTTTGAGATT	9060
15	ATATATATTA ACATTGATGT TCTTTAGTGC CAATGCAATC TTAAACGTGT TTATACCTTT	9120
	ACGAGGGCAT GATTTAGGCG CAACGAATAC GGTTATCGGT ATCGTTATGG GGCATACAT	9180
	GTAAACAGCA ATGGTATTTT GACCATGGGC AGGACAAATT ATTGCTCGTG TCGGTCCCAT	9240
20	TAAAGTATTA AGAATTATTT TGATTATCAA TGCCATAGCT TTAATTATTT ATGGTTTTAC	9300
	TGGCTTAGAA GGTATTTCG TAGCACGTGT TATGCAAGGT GTGTGTACGG CATTCTTTTC	9360
	TATGTCTTTA CAGCTAGGTA TTATTGATGC ATTACCAGAG GAACATCGTT CTGAAGGTGT	9420
25	ATCATGTGAC TCGCTATTTT CAACGATTCC AAACCTAATC GGACCATTAG TTGCCGTAGG	9480
	TATTTGGAAT GCAAATAATA TTCACTATT TGCAATTGTC ATTATCTTTA TCGCATTAAAC	9540
	AACAACATTC TTGSTATCG CGTGACCTTT GCTGAACAGG AACCCGATAC GTCAGATAAG	9600
30	ATTGAAAAAA TGCCGTTTAA CGCTGTAAC TTTTTCGCGC AATTTTCAA AAATAAAGAG	9660
	TTGTTGAACA GTGGTATTAT CATGATTGTT GCATCGATTG TATTTGGTGC AGTTAGTACA	9720
	TTTGTAACGT TATACACAGT GAGTTTAGGA TTCGCGAATG CGGGAATCTT TTTGACAATA	9780
35	CAGGCCATCG CAGTTGTTGC GGCAAGATTT TACTTAAGGA AATACATTCC GTCAGATGGT	9840
	ATGTGGCATC CTAAATATAT GGTATCTGTA CTATCATTAT TAGTAATCGC GTCATTTGTA	9900
	GTGGCATTG GTCCGCAAGT AGGTGCAATT ATTTTCTATG GTAGTGCAT ATTAATAGGA	9960
40	ATGACGCAAG CAATGGTGTA CCCAACATTA ACATCATACT TAAGCTTCGT CTTACCAAAA	10020
	GTAGGTCGTA ATATGTTGTT AGGTTTATTT ATTGCCTGTG CAGACTTAGG TATATCGTTA	10080
45	GGTGGCGCAT TGATGGGACC TATTTCCGAT TTAGTAGGAT TTAAATGGAT GTATCTAATT	10140
	TGTGGTATGT TAGTCATTGT AATAATGATT ATGAGTTTCT TGAAAAAGCC AACACCACGT	10200
	CCAGCGAGTA GTCTTTAATG AAGTGAATTA AAGCATATTA AGTTAATGAA TATTTAAATT	10260
50	TTAAAAGGTA TATTGAGCAT GGCGATTCAT GTGCTTCATG CTAGGACATG AAACATTCTA	10320
	TATGGCTCGT TTTTAGAACG ACATATATCT AAATAAAGCA CGCTTAAG TGAGTTTGA	10380
55		

	TTACATGAAA ATATGCAAAA CGAGTATAAC TGCTAATTGA TAGAAATAGC TCACCATAAA	10500
	ATTACGGTAT GATTTTAAAT ATAAGTAAGT CGCACTACCT GCTAGTATCA ATGCTGGAAT	10560
5	GAATTCCCAC CATGTATTAA TGTATGGATA GTAGAACAGA GTTTC AAGGA TAATGGACAA	10620
	TACTATTGTA ATCTTTAAAG GTATTAATCT GCTTAATTCT TGAATTAAAA TATGACGGAA	10680
	AATAAGTTGA CAAATCAAAG TATTTAATAT AATGGTTAAC GAAAATATAG CTATTAAACT	10740
10	GATGGAaCCA TACCCTTTAA TGAGCGGGTA AATGTCAAAG ACAGTAAAGG AATCTACATT	10800
	TAGTGCgAAA ATATTGAAAT GATTTAAAAG TAAAAAGAGT ACCACACTTA GTGTAAATCA	10860
	TATAAGAATA TGCCATTTAT ATTTAGCACT AGCAACGATT TCGGAACGTA TCATTGGAAT	10920
15	AAACGCATCT TCATGCATCA GACGAAAAAT AGCTAGTGAA ATAATAACTG CGAGTAAATA	10980
	GCTAATGTTC ATTGAAATAG GAAAAGAGAA ACCCCACGGA GCTTGTGAG TGAATACAGC	11040
20	TACTAACCCA AAAGTTAAAA AGACGATAAT GATCGGCAAG ATGTTAACCA AAAATATGTA	11100
	AAGGAAAATA AATCCAATAT CACGTTTGAA AAAACGCGAT TGTTCGGTAG CGTATTCTTC	11160
	TTCTATGTAA TGTTTATTTG TATTTGACAT AGTATACCTC TTAAATAGTT GTATTATATA	11220
25	GATACTTTAG CACATATTAC TTTGTATTGT ATGTTTTATA CATTAAAATT TAAAATGAAA	11280
	AACATATCAT AAAATTGTTT TATAAAATGA AGCGCTTCCA TTGTGTTTTG TTTGTAAAGG	11340
	TGTATCATAA ATATTGAATT GAAATTTTGG GGGGAGGTAT TGTAATGACG TTTCTTACAG	11400
30	TCATGCAATT TATAGTTAAC ATTATCGTTG TAGGATTCAT GCTTACGGTT ATTGTTATCG	11460
	GGCTTATTTG GTTAATTTAA GATAAAAGAC AATCACAACA TAGTGTATTA AGGAATTATC	11520
	CTTTACTAGC ACGTATTAGA TATATTTTCA GAAAAATGGG ACCGGAATTA CGTCAGTATT	11580
35	TATTTTCTGG GGATAATGAA GGGAAACCTT TTTCACGTAA TGATTATAAA AATATCGTTT	11640
	TGGCTGGAAA ATATAACTCT CGTATGACCA GCTTCGGTAC TACTAAAGAT TATCAAGACG	11700
	GCTTTTACAT ACAGAACACA ATGTTTCCGA TGCAACGTAA TGAGATTTCA GTAGATAATA	11760
40	CAACATTGTT ATCAACATTC ATTTATAAAA TCGCGAATGA GCGTTTATTT AGTCGTGAAG	11820
	AATATCGTGT GCCGACAAAG ATTGATCCGT ATTACTTAAG TGATGACCAT GCAATAAAAT	11880
45	TAGGTGAACA TTTAAAACAT CCATTTATTT TAAAACGTAT CGTAGGACAA TCTGGTATGA	11940
	GTTATGGCGC TTTAGGAAAA AATGCCATTA CAGCTTTATC TAAAGGTCTA GCTAAAGCGG	12000
	GCACTTGGAT GAATACAGGT GAAGGTGGCT TATCAGAATA TCATTTAAAA GGTAATGGGG	12060
50	ATATCATTTT CCAAATTGGT CCCGGTTTAT TTGGTGTTCG TGATAAAGAA GCTAATTTTA	12120
	GTGAAGGTTT ATTTAAAGAG GTTGACAGT TATCTAACGT ACGCGCATTT GAGCTGAAGT	12180

55



	TTGCTAAAAT CCGAAATGTT GAACCTTATA AAACAATCAA TTCACCTAAC CGTTACGAAT	12300
	TTATTCATAA TGCTGAAGAT TTGATTGCTT TCGTTCGATCA GTTGCAGCAA TTAGGTCAAA	12360
5	AACCAAGTAGG ATTCAAAATT GTAGTAAGCA AAGTTTCAGA AATTGAAACA CTTGTACGTA	12420
	CGATGGTGGG ACTAGATAAG TATCCAAGCT TTATTACGAT TGATGGTGGT GAAGGTGGTA	12480
	CTGGTGCAC ATTCCAAGAA TTACAAGATG GTGTTGGCTT ACCGCTATTT ACAGCTCTAC	12540
10	CTATTGTGTC TGGCATGTTA GAAAAATATG GTATTGAGA TAAAGTGAAA TTGGCGGCAT	12600
	CTGGTAAGTT AGTGACACCA GATAAAATTG CGATTGCACT AGGTTTAGGT GCAGATTTTG	12660
15	TAAATATCGC ACGTGGGATG ATGATTAGTG TCGTTGTAT AATGAGTCAA CAATGTCACA	12720
	TGAATACGTG TCCTGTAGGT GTTGCAACGA CAGATGCGAA GAAAGAAAA GCATTGATTG	12780
	TTGGAGAAAA GCAATATCGT GTCACAACT ATGTAACAAG TTTGCATGAA GGCTTATTCA	12840
20	ATATTGCAGC AGCTGTTGGC GTATCCAGTC CTACAGAAAT TACTGCTGAT CATATTGTAT	12900
	ATCGAAAAGT CGATGGTGAG TTACAAACGA TACATGATTA TAAATTAAAA CTCATTAGTT	12960
	AACTTAATTA TTTCCGGAAA TTGAAAGCAG CGGATTTTAG CGTTACTGCA AATAATTTTA	13020
25	TATTAGTAGT GGATGCTGGT CACACAAGAA CTTCAAATAT TAAAGCCCTC AGAATATGAA	13080
	TTAAGGTTTG TAACCTTAGT CTTATCTGAG GGCATTTTAA AGTTATAAAC TATTGTGCT	13140
	CCATTTTATC TTTTCTTTT AAACCTCTGT GCTTTAATTG CTTTCAAGT TTTTCAAAAC	13200
30	TAATATCTTT ATTTCTTTA GTCGAAACAC CAAGACGTTT ATTTAATTTT TTCATGTCAA	13260
	CTTCTGTGTA ATCTATGTCT AAGTGyTCAA TTGCTTTTTT ATCTTTATAG TCTACTTTGT	13320
	ATTTTACGCC TTTAAGGTCT TTGAAAATAC TTTTCAGATTT GGCGAATAAC TTTTGGCTT	13380
35	CGTCTTTATC CATACTAGA TCGTCATATT TAATTGTGTT GATTGTAGAC TGTTTTAAAA	13440
	CTTTATCATC TTTATATGTG ATAGAAGTTA GTACATGTTT ACCACTAACA TCACCWTCAT	13500
	ATGTTTTGGT TTGTTCTTTA CCACAAGCTG ATAATGCAAT GATACAACT AATGCTACTA	13560
40	CAATTAATGA ACATAATTTT TTCAAAGTCA GTCGCCTTCT TTCGATATTT GTATTATAAA	13620
	GAAATTATAA CATTTACTAA AAAATGATGT TATTCAAAAA TTAAATTTT GTCATTTTTT	13680
45	TTGAAGATAT GAGTTTTTTT AAGCGGATTC CTCACAAAAT TTAAAAATA TTAAAGCCTk	13740
	AAAATGATAA AGCGkTAGGG AACGTTTTTC TGAAAGTTAG TGATACAATA GTTTTAAGTT	13800
	GAAATACAGG AGGATGAATA ACATGAATCA GTCAGTCAAA TTAATAAAC ATTTAACAGA	13860
50	TGTAAACGGC ATTGCTGGTT ATGAAATGCA AGTTAAAGAA GCAATGCGTa ACTATATAGA	13920
	GCCTGTCAST GATCAAAITA TTGAAGATAA CTTGGGTGGC ATTTTGGAA AGAAAAATGC	13980

55

	AACAAAGATT GATAAACATG GTTTTATTTT ATTTACGCCA kTgGTGGATG GTGGAATCAA	14100
	GTCAATGCTAT CTCAAAAAGT AACGATTACA ACAGATTCGG GCAAAGAAAT TAGAGGTATC	14160
5	ATCGGTTCTA AACCGCCACA TGTCTTAACG CCTGAAGAAC GTAAAAAGCC AATGGAAATC	14220
	AAAAATATGT TTATAGATAT TGGTGTAGT AGCAACGAAG AAGCTGAAGA AGCTGGCGTT	14280
	GAAGTAGGCA ATATGGTTAC GCCATATAGT GAATTTGAAG TGCTTGCAAA TGATAAATAT	14340
10	TTAACTGCCA ATCATTGAT AATCGCTATG GCTGTGCATT AGCTATTGAG GTATTAAAC	14400
	GTTTAAAGA TGAAATATT GGCATTAAC TATACAGTGG TGCCACAGTG CAAGAAGAAG	14460
15	TGGTTTTCG TGGTGGCAA GTGGCAGCGA ATACGATTAA ACCAGACTTG GCGATAgcTG	14520
	TcGATGTAGG TATTGCTTAT GATACCCAG GTATGTCAGG TCAAACGAGC GATAGTAAAC	14580
	TAGGCGGTGG TCCAGTTGTC ATTATGATGG ATGCTACAAG TATTGCTCAC CAAGGTTTGC	14640
20	GAAAgcATaT TAAAGATGTA GCTAAGGAAC ATAACATCGA AGTACAATGG GATACGACAC	14700
	CAGGTGGAGG TACAGATGCG GGAAGTATTC ATGTCGCAAA TGAAGGTATT CCAACGATGA	14760
	CAATCGGTGT TACGCTGCCA TACATGCATT CTAATGTTTC AGTGCTCAAT GTAGATGATT	14820
25	ATGAAAAATC TATCGTCTT GTTACTGAAA TTGTCCGTTT ATTGAATGAT GAAAGTTATA	14880
	AAAATATCAT GTGGTAATCA AATCCATAAA TAATAAGAA TCCTTTTAAT ATGGTAGGTT	14940
	GTAAACAAT TGTCTAATTT TAATTCCTAG TCATTAGACA GTATCCATGT TAATAGGATT	15000
30	TTTTGTTTTT AATTTAAATG CTGAAAATCA ATTATGCCA AATTTTGATA TTACAAGAAA	15060
	ATGATTTTTT CTAAATGTA ATTGCACTAA AAACCAAAAA AACGGGAATA ATATACCTGA	15120
	TATATTACAT GAGGAGCGGT GCAAATGTTG TTAGAAATTA AAGATTTAGT GTATAAAGCG	15180
35	AGCGATAGAA TCATACTAGA TCATATCAGT CTAAAAGTAG ATAAAGGCGA GAGTATTGCC	15240
	ATTATAGGTC CATCAGGTAG TGGTAAAAGT ACATTTCAAA AGCAAATATG TAATTTGTTT	15300
	AGTCCAAC TA GTGGAGAACT TTATTTTAAA GGTAAACCCT ATAATGATTA TGACCCGGAA	15360
40	GAATTGCGTC AACGAATCAG TTATTTGATG CAGCAAAGTG ACTTGTTTGG TGAAACGATT	15420
	GAAGATAACA TGATATTCCC ATCACTTGCA CGTAATGATA AATTTGATAG AAAACGTGCA	15480
45	AAGCAATTAA TTAAAGATGT CGGTTTGGGA CATTATCAAT TAAGTTCGGA AGTGGAAAA	15540
	ATGTCGGGTG GTGAGCGGCA AAGAATTGCT ATAGCGCGCC AACTGATGTA TACACCGGAT	15600
	ATTCTTTTAT TAGATGAATC GACCAGTGCA TTAGACGTTA ATAATAAGA AAAGATAGAA	15660
50	AATATCATTT TTAAATTAGC AGATCAAGGC GTGGCAATTA TGTGGATTAC CCACAGCGAT	15720
	GACCAAAGTA TGCGACACTT TCAAAAGCGT ATAACAATTG TTGATGGTCA AATTTCTAAT	15780
55		